Simulation of Traveling-Wave Tubes for Analysis and Optimization in Modulated Back-Off

Vom Promotionsausschuss der
Technischen Universität Hamburg
zur Erlangung des akademischen Grades
Doktor-Ingenieur (Dr.-Ing.)

genehmigte Dissertation

von
Djamschid Safi

aus
Hamburg

2020

First Examiner:	Prof. Dr.-Ing. Arne F. Jacob
Second Examiner:	Prof. Dr. rer. nat. Dr. h. c. Manfred Thumm
Chair of Examination Board:	Prof. Dr.-Ing. Gerhard Bauch

Date of Thesis Defense:	June 22, 2020

Acknowledgment

In the following, I would like to take the opportunity to thank some of the many people who made this work possible by their support.

First, I would like to thank Prof. Dr.-Ing. Arne F. Jacob, who gave me the opportunity to explore an exciting and challenging field of study and has always supported me both professionally and personally. Also, I would like to thank Prof. Dr. rer. nat. Dr. h. c. Manfred Thumm for taking the time to prepare the second expertise and Prof. Dr.-Ing. Gerhard Bauch for chairing the examination committee.

This work was performed in collaboration with the company Thales. Here, Dr.-Ing. Philip Birtel in particular was not only an ally and colleague, but also a teacher, without whom this work would not have been possible. I would also like express my gratitude to Dr. Jürgen Wegener and Monsieur Jean-François David for all these fruitful discussions at Thales.

Likewise, I would like to address a thank you to my colleagues and thematic comrades-at-arms at the Hamburg University of Technology, especially Moritz Hägermann for the regular exchange and thorough proofreading, and Dr.-Ing. Sascha Meyne for being a fantastic mentor on top of that.

But, of course, I would also like to thank the students who accompanied me during my doctorate and whose dedication has contributed much to this work: Till Pusch, Lukas Buderath, and Jonathan Ritter within their bachelor's theses and Michael Wulff for his ideas and efforts as a student assistant and within his bachelor's thesis.

I cannot forget to thank my friends and my family, who supported me at all times.

And, most importantly, thank you, Louise. I simply could not have done any of this without you being there.

Contents

1 Introduction

Traveling-wave tubes (TWTs) are high-power, high-efficiency vacuum electronic amplifiers for radio-frequency (RF) applications. Besides their predominant application in satellite communications, they also play an important role in airborne and surface radars, as well as in electronic countermeasure applications. In TWTs, a propagating RF wave is amplified by extracting kinetic energy from a synchronous electron beam, reaching average RF powers in the range of watts to hundreds of kilowatts. Within the group of microwave vacuum electronic devices, TWTs are the commercially second most successful type[1] [1], [2], as for many applications they are unmatched in terms of combined power, bandwidth, and cost.

1.1 Traveling-Wave Tubes in Modern Satellite Communications

Power amplifiers in satellite transponders need to fulfill a number of requirements, such as high output power, high efficiency, large bandwidth, low mass, and long lifetime of more than 15 years [3]. In most of these terms, today's TWTs are hard to beat. At C-Band, for instance, the available output power of commercial devices grew by a factor of 40 compared to the first commercially used TWTs, while the weight was reduced by a factor of 2-6 [3], [4]. At the same time, research on TWTs has resulted in efficiencies starting from as little as 1 % to nowadays above 75 % at saturation [4].

This development has led to a continuing dominance of TWTs in the satellite sector compared to solid-state power amplifiers (SSPAs). Considering, for instance, the failure rate per 1 W as of 2004, TWTs have shown to be more than six times [5] up to one order of magnitude more reliable [2]. While newer studies suggest that this gap is slowly closing [6], [7], the data indicates that it will remain significant within the foreseeable future. Also, the continuously improved performance of TWTs resulted in a much smaller overall cost compared to solid-state devices, as impressively shown by the numbers from 2003 given in [2]. There, the findings of a "major satellite manufacturer"

[1] Number one is the microwave oven's magnetron.

are recited, stating they would require three times the number of satellites, resulting in an increased cost of $ 1 billion when using SSPAs instead of TWTs.

This trend was enabled by constant improvement in two main areas. The rise of multi-stage depressed collectors in the 50's and 60's up to the first three-stage depressed collector in the 70's led to a significant reduction of losses generated in the collector. From there on, in parallel to further collector development, the beam efficiency has rapidly been increased by delay-line velocity tapering. Nowadays, optimized TWTs utilize four- or even five-stage depressed collectors and complex delay-line tapers.

In recent years, with the introduction of the DVB-S2 and the latest DVB-S2x standards [8], [9], the operating mode of TWTs in such satellite links has changed. In these modern applications, the modulation schemes shift from simple ones, like constant-envelope quadrature phase-shift keying (QPSK), to more involved amplitude- and phase-modulated signals. This is important, as – like any physical amplifier – the TWT introduces strong nonlinear distortions at high power levels according to the signal envelope. To avoid these distortions, the TWT typically is operated far in the linear region below saturation, i.e., in back-off. While this diminishes nonlinearity-induced modulation errors, it severely reduces the overall efficiency, as the beam efficiency decreases quickly. As these more complex modulation schemes typically require better linearity, this issue calls for increased attention [10].

Currently, when simulating TWTs, their performance is mainly characterized in the near-saturation operating state based on their narrowband properties, such as the output power or the nonlinear phase shift. Specialized TWT simulation tools, such as MVTRAD [11] or CHRISTINE [12], are typically designed to calculate these frequency-domain properties for saturation-based optimization. For the aforementioned modern operating modes, this does not suffice anymore. The modulation of the input signal results in continuous, more broadband frequency spectra. The consequential distortions have a strong influence on the transmission quality, but can only rudimentally be predicted by classical steady-state codes. Thus, these classical simulation routines do not directly allow optimization of the TWT customized to modern applications.

This work aims at providing suitable solutions to this problem. The goal is to be able to carry out extensive simulation of the downlink chain of a satellite, operated with different modulation types. This requires the development of methods, which enable the fast and accurate calculation of meaningful multi-tone and multi-carrier characteristics.

From today's point of view, there are three promising approaches to tackle this specific task. General purpose simulation tools, such as CST Particle Studio's [13] Particle-In-Cell (PIC) [2] solver or MAGIC [14], operate directly in time-domain and inherently allow the simulation of arbitrary input signals. Thus, they can be used to study, e.g., transients, memory effects, and stability issues. Unfortunately, this comes at the cost of a high computational effort and a fairly involved modeling, which is why their reliability requires thorough investigation. Frequency-domain codes, on the other hand, are well-established for the simulation of TWTs and are known to accurately predict the nonlinear interaction. As they rely on a number of simplifying assumptions they are able to significantly reduce the computation times for single-tone excitation compared to general purpose tools. The calculation of meaningful multi-tone characteristics, on the other hand, demands for multi-tone input signals. This can be achieved using a large number of harmonics, which requires adjusting the numerical parameters. As a consequence, this strongly increases the numerical effort. A third option is the usage of so-called Envelope methods. Assuming a slowly varying modulation of the TWT's excitation, these methods enable a fast calculation of the output signal in time-domain, based on data provided, e.g., by a fast frequency-domain code. Simple Envelope approaches have proven to be accurate for narrow carrier spacings, while their limits for TWTs driven by signals with complex modulation schemes are not clear a priori.

These three methods are complementary to each other. In this work, suitable modeling approaches are developed and characterized regarding their computational effort, their practicability, and the accuracy of the results, which is additionally verified by measurement data. Suitable approaches enabling the analysis and optimization of TWTs for modern operating modes need to be as fast as possible to allow their usage, e.g., at the design stage, but also as accurate as necessary to yield reliable performance indicators. The development of such methods is undoubtedly crucial to further improve TWTs in back-off and ensure their competitiveness in future applications.

1.2 Outline of this Thesis

This work attempts to lay groundwork for tackling the challenges of using and improving TWTs for modern communication satellite systems. For this purpose, an essential component is the ability to obtain common communication characteristics of the TWT with as little effort as possible, at best already at the design stage. Thus, after a brief introduction to important basics of TWTs in Chapter 2, the following Chapter 3 covers various simulation methods to predict the transmission behavior of TWTs for such modulation schemes. This is a prerequisite for understanding the effect of nonlinearities on communication signals and consequently optimizing the TWT regarding its actual operating conditions, as is discussed in Chapter 4.

For simplicity, comparability, and consistency, most of the investigations in this work are carried out on a state-of-the-art 150 W commercial reference TWT for communication in Ku-Band. In some cases, a distorted version of it is used to highlight certain frequency-dependent effects, resulting in a more dispersive TWT simulation model. Results from this "crooked" TWT are tagged accordingly whenever they are considered. In addition, for validation purposes, Section 3.4.4 features the simulation of a folded-waveguide (FW) TWT as a representative of future technology in the field of satellite communications.

2 Basics of Traveling-Wave Tubes

2.1 Components

Accurate modeling, simulation, and optimization of TWTs require a solid understanding of the main elements playing a role in the amplification process. Therefore, these components will briefly be described in the following, starting from the electron gun up to the collector. In Figure 2.1, a simplified schematic of a helix TWT is shown, indicating all covered components.

2.1.1 Electron Gun, Beam, and Magnetic Focusing

The electron beam, which is required for the interaction with the traveling wave, is generated in an electron gun. For this purpose, a cathode is heated by a filament at typical temperatures of around 1000 K to 1500 K, such that electrons are emitted following the thermionic emission mechanism [1], [3]. Lifetime and current density considerations lead to a large, spherical cathode. The cathode is followed by a focus electrode, which serves to bundle the beam towards the desired beam radius. One or several anodes are used to further shape the beam, control the current flow, and accelerate the electrons according to the applied anode voltages. Depending on the application, the velocity of the electrons leaving the electron gun reaches around

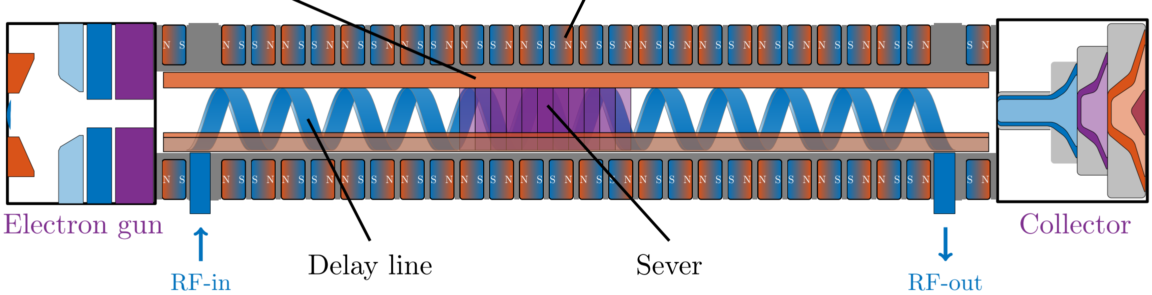


Figure 2.1: Schematic view of a helix TWT.

$10 - 40\,\%$ of the speed of light. The electron beam is typically focused by a periodic permanent magnet (PPM) configuration, since it allows a weight reduction of up to two orders of magnitude compared to solenoidal magnets [1]. This is especially important for satellite applications. According to Busch's theorem [1], a magnetic flux B_c through the cathode causes a rotation of the beam in azimuthal direction θ, with angular velocity

$$\frac{\partial \theta}{\partial t} = \frac{\eta}{2}\left(B - B_\mathrm{c}\frac{r_\mathrm{c}^2}{r_\mathrm{b}^2}\right), \tag{2.1}$$

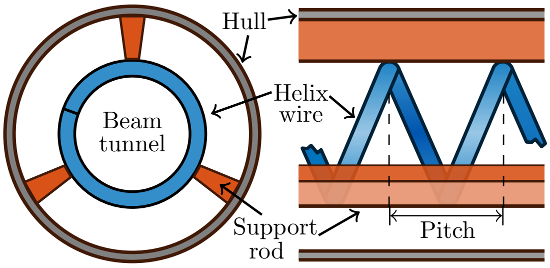
where r_c and r_b are the cathode and beam radius, respectively, and η stands for the electron charge to mass ratio. B denotes the axial magnetic field, which according to the Lorentz force radially focuses the rotating beam, counteracting the space charge forces and their diverging effect on the beam.

2.1.2 Components of the Interaction Region

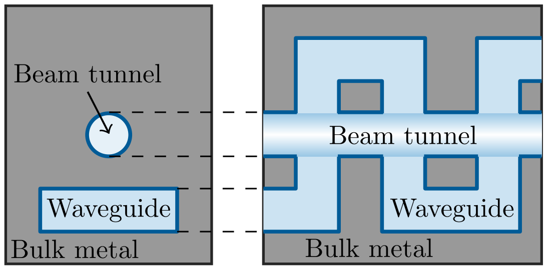
The interaction region denotes the entire area between electron gun and collector, where electron beam and traveling wave are coupled for the sake of energy transfer. The principle behind this energy transfer is outlined in Section 2.2. The main ingredient of the interaction region is the so-called delay line or slow-wave circuit. The amplifier's performance and properties strongly rely on the type of delay line and its implementation. The main purpose of the line is to slow down the electromagnetic traveling-wave in axial direction by forcing it to take a defined detour along the slow-wave structure. This is essential to enable coupling between the electron beam and the traveling wave. Significant interaction occurs, if the axial phase velocity of the wave traveling on the delay line is close to the velocity of the electrons in the beam. There are various types of delay lines with different fields of application, the commercially most important ones to be mentioned are the helix and the coupled-cavity line.

Helix TWTs offer very large bandwidths of up to several octaves, at the cost of a comparatively moderate power capability. The circuit, as it is shown in Figure 2.2a, consists of a wire which is wound up on a helical path around the beam axis. Especially the pitch and the radius of the helix determine its properties, serving as set-screws to adjust the phase velocity and the coupling to the beam. The helix wire is fixed inside the TWT's hull by means of ceramic support rods and therefore exhibits poor thermal properties. Helical delay-lines are the dominant topology for satellite communication applications at the currently most relevant frequency bands.

For high-power applications, such as Radar, coupled cavities out of bulk metal are advantageous. There, the electron beam travels through a beam tunnel and interacts



(a) Helix.



(b) Folded-waveguide.

Figure 2.2: Two suitable delay-line types for satellite communications.

in gaps along the line with the electromagnetic wave in the cavities. While their amplification is rather narrowband, their power capability is significantly larger due to the thermal conductivity of the bulk metal. Somewhere in between these delay-line types lies the FW geometry, shown in Figure 2.2b. It can be seen as a compromise between the broadband helix and high-power coupled-cavity lines and is currently investigated as a promising alternative for communication satellite applications at Q-Band [15] and above. A rectangular waveguide is folded around a beam tunnel to enable the required phase synchronicity. The winding can either be smooth in a meander-like manner or with sharp 90°-corners, as depicted in Figure 2.2b. Like for coupled-cavity delay-lines, the beam interacts with the traveling wave within the discrete gaps at the crossing points between the beam tunnel and the folded waveguide.

Regardless of the delay-line type, typical TWTs exhibit high amplification. Today's helix TWTs can easily reach 60 dB in gain and more. Thus, stability against oscillation, for instance due to mismatched delay-line couplers at the RF input and output, has to be considered. For this purpose, one or several severs are introduced. They are strongly attenuating and well-matched elements, which in practice, for instance, are realized by means of thin layers of lossy material on the support rods. Effectively, severs separate the interaction region into several sections, such that the loop gain conditions for avoiding undesired oscillations is met everywhere. While the power of the traveling wave is fully absorbed in the attenuator, the modulated electron beam drifts through it and regenerates the signal afterwards.

The bandwidth and frequency dependency of the individual TWT types is a result of the delay line's dispersive property. According to Floquet's theorem [16], the electromagnetic field has the same periodicity as the delay-line except for a phase factor φ. Thus, it is reproduced along the periodic structure with this phase factor from one unit cell to another. This is often visualized by means of the dispersion diagram, which describes the phase advance on the delay line along one unit cell pitch p versus frequency. In simulation, the diagram can be obtained by means of an eigenmode analysis, where the unit cell of the delay line is modeled with phase-shifted periodic boundary conditions in axial direction. Thus, a propagation constant

$$\beta = \frac{\varphi}{p} \tag{2.2}$$

can be obtained for the calculated eigenmode.

Due to the periodicity, the axial electric field can be expanded into a Fourier series [17]

$$E_z(r, \theta, z) = \sum_n E_n I_n(\gamma_n r) e^{-j\beta_n z} e^{-jn\theta}, \tag{2.3}$$

where I_n denotes the modified Bessel function of n-th order and first kind, with one phase propagation constant

$$\beta_n = \frac{\varphi}{p} + \frac{2n\pi}{p} = \beta_0 + \frac{2n\pi}{p}, \tag{2.4}$$

and one phase velocity

$$v_{\mathrm{ph}} = \frac{\omega}{\beta_n} \tag{2.5}$$

for each space harmonic n, but a common group velocity

$$v_{\mathrm{gr}} = \frac{\partial\omega}{\partial\beta_0}. \tag{2.6}$$

For a known field shape, each space harmonic with index n can be described by a single scalar E_n, which is extracted from the eigenmode calculation. Typically, instead of E_n, the coupling impedance [17]

$$Z_{\mathrm{c},n} = \frac{|E_n|^2}{2\beta_n^2 P_{\mathrm{t}}} \tag{2.7}$$

is used to characterize the coupling strength, with P_{t} describing the total transported RF power.

2.1.3 Multi-Stage Depressed Collector

A main advantage of TWTs compared to SSPAs is their high total efficiency. A significant contributor to this figure of merit is the collector, where the spent beam[1]

[1] The beam is called "spent", when it has passed through the interaction region and the DC input power has been used for interaction. Thus, the term denotes the beam at the interface between delay line and collector. It is considered as spent, no matter which part of its energy is transferred.

at the end of the interaction region is collected to reuse as much of the remaining beam power as possible. For this purpose, the focusing is removed to enable the beam to diverge. The collector potential is depressed from the TWT's body potential. Otherwise, the electrons would hit the collector's walls at full speed and their kinetic energy would largely be converted to heat. A potential depression, where, relatively to the cathode, a reduced voltage V_{Coll} is supplied to the collector, leads to a collector current I_{Coll}, which provides more power from the beam to the DC power supply, while the same current is processed. Therefore one can also consider this a DC input power reduction by

$$P_{\mathrm{DC,out}} = V_{\mathrm{Coll}} I_{\mathrm{Coll}}, \tag{2.8}$$

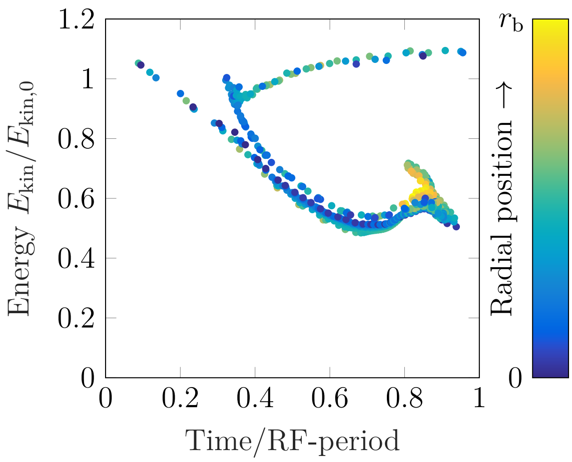
rather than a DC power recovering.

The electrons of the spent beam entering the collector exhibit a variety of velocities, depending on, e.g., the delay-line taper, frequency content, and drive level. The spent beam's velocity spread is quantified in the electron-velocity spectrum (EVS). For visualization purposes, the electrons are sorted according to their kinetic energy and accumulated up to the total current entering the collector. Figure 2.3a shows an exemplary spent beam resulting from a frequency-domain simulation for a large-signal single-tone excitation close to saturation. The color represents the radial positions of the particles, with yellow being those at the outer radii and dark blue the inner ones. Due to its steady-state background, the resulting temporal beam distribution is periodic according to the fundamental frequency. In Figure 2.3b, the shown period is evaluated with respect to the kinetic distribution of the current and compared to a small-signal result and the unmodulated case. The energy axes are normalized to the initial kinetic energy $E_{\mathrm{kin},0}$, and the current to the total beam current I_0. A strong variation of the spectral properties with drive is apparent.

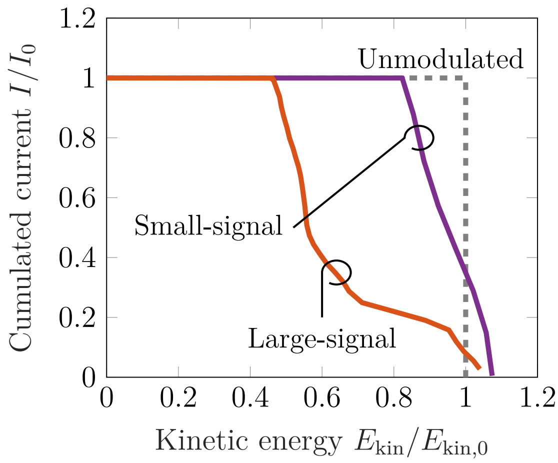
As the beam is slowed down by the collector depression, the maximum power that a single-stage collector can recover is limited by the slowest expected electrons, arriving at the collector with kinetic energy $E_{\mathrm{kin}}^{\mathrm{min}}$. The amount by which the potential of the collector V_{c} is depressed against the helix potential V_{Helix} must not fall below the voltage at which the slowest electrons cannot reach the collector anymore. Thus, it is limited by

$$q\,\left(V_{\mathrm{Helix}} - V_{\mathrm{c}}\right) < E_{\mathrm{kin}}^{\mathrm{min}}, \tag{2.9}$$

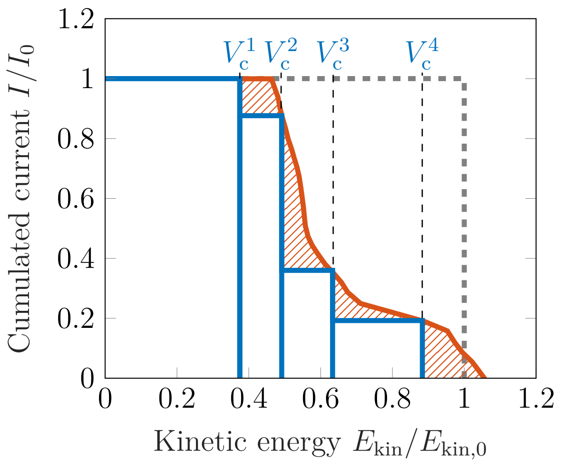
with the elementary electron charge q. A breach of this condition would allow the electrons to be reflected back into the interaction region and disturb the amplification or even damage the device. To cover as much of the electron velocity distribution of the spent beam as possible, modern collectors use several stages at different depressed



(a) Spent-beam energy in one RF period.



(b) EVS from spent beams.



(c) Energy recovering in the collector.

Figure 2.3: Large-signal spent beam, the corresponding EVS, and the recollection in a multi-stage depressed collector.

potentials, resulting in a DC power reduction of

$$P_{\mathrm{DC,out}} = \sum_i V_\mathrm{c}^i I_\mathrm{c}^i, \tag{2.10}$$

where a collector current I_c^i is generated in the i-th collector stage at potential V_c^i. Figure 2.3c shows the spent beam EVS from a large-signal simulation and how it is collected in a perfectly sorting multi-stage depressed collector with stages $V_\mathrm{c}^i, i = 1, \ldots, 4$. For each voltage, one can mark a point on the EVS curve, which determines the collected current on the stage. Then, the sum of the four rectangular areas framed in blue determines the power reduction. In consequence, the patterned red area describes the power that is dissipated. Thus, the energy distribution of the spent beam has a large influence on the possible DC power reduction, being optimal when the EVS appears stair-like, according to the stage voltages.

2.2 Operation Principle

For the operation principle of TWTs, various figurative approaches can be found in literature, such as the surfer, whose velocity has to match to the (water) wave's phase velocity for effective energy transfer[2], or a stream of cars moving over a hilly street [3], where on the uphill side the traffic is decelerated forming a bunch towards the hilltop and the opposite occurs on the downhill side. The need for such simplified and more conceptual explanations hints at the general complexity behind the functional principle of TWTs. A basic description of the helix TWT's operation, as for instance given in [1], is outlined in the following.

Applying an RF signal on the helical line, a charge distribution according to the sinusoidal field distribution of the signal frequency can be observed on the helix, moving along the line at a certain phase velocity. This can be seen in Figure 2.4. In the shown configuration, one wavelength equals four turns of the delay line. A force on the electrons in the electron beam is found following the electric field lines, which are more concentrated inside the helix. Assuming an initial velocity from left to right, alternating accelerating and decelerating areas can be determined. Thus, an accelerating force acts on parts of the beam and a decelerating one on others, leading to areas of bunched particles, moving from left to right. When the phase velocity of the traveling wave is slightly smaller than the velocity of the beam, the electron bunches move even further into the decelerating region.

[2] Andrei Haeff, one of the inventors of the TWT, supposedly got the idea watching surfers, coining the term "traveling wave". [18]

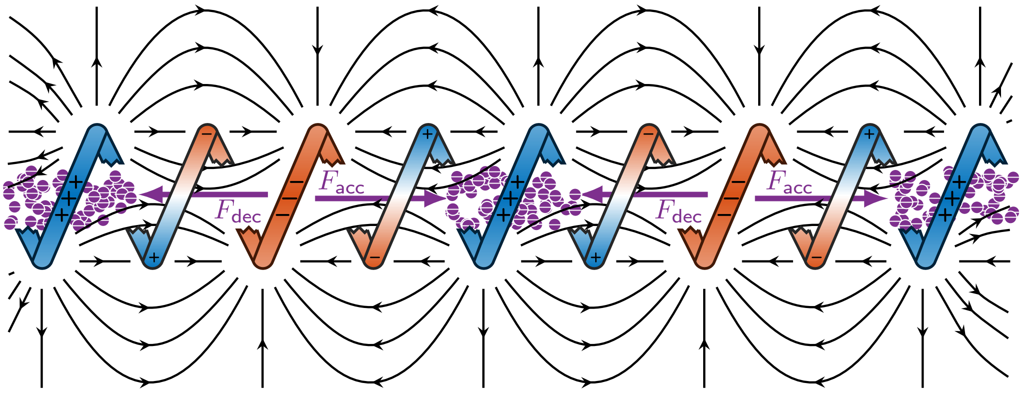


Figure 2.4: Charges on the circuit and resulting electric field for an RF signal applied to a helical delay line. F_{acc} and F_{dec} denote accelerating and decelerating forces acting on the electrons, which, for reasons of clarity, are downscaled in numbers and indicated by purple dots.

The bunches represent charge accumulations, which themselves act on the charges on the helix wire. These charges on the line are pushed away from the bunches, leading to a generally sharper charge distribution and thus an enhanced field on the helix. As the charge on the helix moves away from the bunch in both directions, its velocity to the left and the right of the bunch differs relatively to the beam's DC velocity. As a result, the electrons in front of the bunch accumulate, such that a phase shift of 90° occurs. The electron bunches are thus further decelerated and the bunches on the beam and the accumulations on the line increase, leading to further amplification of the signal.

This qualitative explanation is still fairly simplified and does not enable the quantitative characterization of the process. For this purpose, a small-signal theory of the fundamental principle was developed by Pierce [17]. There, an equivalent slow-wave circuit is considered, interfaced with the closely passing beam which serves as a distributed current source to the circuit. In the synchronous case, where the beam velocity equals the traveling-wave's phase velocity, a determinantal equation with four solutions for the different waves propagating on the circuit can be found. These have different propagation characteristics, three traveling in forward direction and one in backward direction. Of the forward waves, one shows exponentially growing behavior. From this, the linear small-signal gain G for a homogeneous section can be estimated to

$$G \approx -9.54 + 47.3\,CN\,\text{dB}, \tag{2.11}$$

where N denotes the length of the section divided by the signal wavelength, and C is the Pierce gain parameter [1], [17]

$$C = \frac{KI_0}{4V_0}, \tag{2.12}$$

with K being the circuit impedance[3], I_0 the beam current, and V_0 the beam voltage. For this estimate, a lossless, synchronous operation is assumed and large-signal effects such as large-signal space charge considerations are neglected. Extended discussions of the TWT's operation theory can be found, e.g., in [1], [17].

2.3 Nonlinearities and Nonlinear Effects

The TWT is, as any physical amplifier, a nonlinear device. This can also be seen in typical dynamic amplifier characteristics, such as the output power, which is schematically shown in Figure 2.5 versus the RF input power of the TWT. The usable small-signal region, where linear gain is observed as described in Section 2.2, is limited at the lower end by the noise generated in the TWT, in which the low-power signal then perishes. Above the noise floor, a change in input power leads to a proportional change in output power, as the gain is constant. At some point, for increasing input power, the gain decreases. This constitutes the beginning of the large-signal region. Drawing from [2], the following shortly outlines the underlying mechanism leading to this loss of linearity, as well as typical means to mitigate it.

2.3.1 Causes of the Nonlinear Behavior

While the delay line itself is linear, the coupled beam-wave system is inherently nonlinear in nature. The electron beam can be described as a superposition of its steady component and a periodic component, which increases in axial direction and describes the beam modulation. We call these the DC and the AC part of the beam. For the sake of simplicity, a steady-state excitation of a homogeneous delay line with frequency ω is considered.

At the injection the beam only consists of a DC part. This changes with interaction, as, due to the coupling, a velocity modulation is impressed on the beam, according to the RF signal on the delay line. Thus, the electron velocity v has both a DC part v_0

[3] In [1] and others, the circuit impedance is denoted by Z_c, sometimes leading to some confusion. The circuit impedance sums up all information required on the circuit, such that no further information on the actual representation is required.

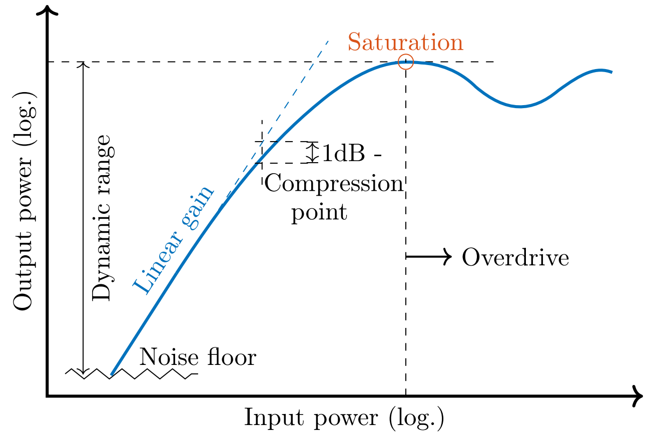


Figure 2.5: Amplifier characteristics of a TWT.

and an AC part v_1, with

$$v = v_0 + v_1 \cdot \exp(j\omega t)\,. \tag{2.13}$$

As a result of this modulation, electrons travel at different velocities and the beam starts to bunch. Consequently, the space-charge density ρ is modulated, too. It also consists of both a DC component ρ_0 and an AC component ρ_1, with

$$\rho = \rho_0 + \rho_1 \cdot \exp(j\omega t)\,. \tag{2.14}$$

As to these descriptions, both AC components of v and ρ at first follow harmonic functions.

For the current density J, their product

$$J = v \cdot \rho \tag{2.15}$$

$$= \underbrace{v_0\rho_0}_{=:J_0} + \underbrace{(v_1\rho_0 + v_0\rho_1 + v_1\rho_1 \cdot \exp(j\omega t))}_{=:J_1(t)} \cdot \exp(j\omega t)\,. \tag{2.16}$$

is considered. The current density already shows harmonics of the drive frequency, and its AC amplitude itself is time-dependent with the RF frequency. Even under purely ballistic conditions, neglecting repelling forces between the particles and determining their movement independently of each other, with time even higher harmonics can occur, as the beam density is further peaked and more sharply bunched.

The harmonics of the beam current in turn effectuate an electric field at the harmonic frequencies on the delay line, as the current J_{ind} induced on the helix relates to the beam current J with [1]

$$\frac{\delta J_{\text{ind}}}{\delta z} \propto \frac{\delta J}{\delta z}, \tag{2.17}$$

which then results in a related electric field. This electric field again has a decelerating and accelerating influence on the modulated beam.

We can consider the beam to be a nonlinear source. The signal on the line now consists of both the fundamental frequency component and harmonics of it, which again cause a – now more involved – velocity modulation and space-charge density modulation. By multiplication, this introduces further harmonics and also intermodulation products at sums and differences of multiples of the mixed frequencies. Some of these generated intermodulation products introduce components at the fundamental frequency. Thus, there, the power is determined by superposition of the initial fields and the intermodulation products generated in the nonlinear beam.

The model of ballistic bunching is of course a simplification, which fits well far in back-off, where the excitation is small and the bunching forces are weak. When considering higher drive levels, the modulating forces and therefore also the mixing becomes stronger. This requires taking space-charge forces into account. As the electrons interact with each other, their repelling forces would, lacking an external force, lead to a diffusion of the bunches. Thus, these can only be sustained, if the delay line contributes with a sufficiently large axial electric field. This traveling wave is amplified taking power from the beam in the coupled process and at the same time sustaining the bunches of the beam.

On average, the kinetic energy of the electron beam is transferred to the electric field, while the kinetic-energy spectrum of the beam is spread. Therefore, the beam does not only slow down on average, but subsequently also loses synchronism with the wave. As this synchronism is required for the traveling wave to counteract the debunching, the velocity modulation spread leads to a relaxation of the space-charge modulation. This effect depends on the drive level and causes the compression of the gain curve. At some point, the bunches cannot be sustained anymore and no more kinetic energy can be transferred from the beam to the wave. This limits the amplification process as the amplifier is in saturation.

At the same time, the reduced average beam velocity influences the phase shift observed in the RF wave. The superposition of the traveling wave with the intermodulation products generated in the nonlinear beam already results in a phase shifted oscillation. The phase shift increases with the amount of extracted power, as the slower beam additionally pulls back the RF wave. Further increase of the RF excitation above saturation even reduces the maximum power at the output, as energy partially is transferred from the traveling wave back to the electron beam.

In the nonlinear region, two main effects occur. The first one is the gain reduction or compression and is often denoted as amplitude-modulation to amplitude-modulation

(AM-AM). The second one relates the phase change which is caused by a modulated amplitude to the phase in small-signal operation. It is often called the nonlinear phase shift or amplitude-modulation to phase-modulation (AM-PM). These two characteristics are effective metrics to describe the nonlinearity and useful for predicting nonlinear distortion or intermodulation, as will be shown in Chapter 3. Still, it is important to note that they are not a cause of intermodulation, but rather caused by it, as discussed in this section. Also, the AM-AM and AM-PM characteristics only summarize what is seen from the outside of the TWT, possibly distorted by the coupler, while the intermodulation itself in principle occurs throughout the device.

2.3.2 Nonlinearity Compensation

In the context of satellite communications, distortions as introduced by driving the TWT in the nonlinear regime are to be avoided. In the following, two approaches to reduce these distortions are briefly outlined.

Predistortion Linearization

Predistortion linearizers [19] are commonly used in satellite communications and therefore an integral part of a holistic downlink chain investigation. They rely on conditioning the input communication signal by means of a preceding nonlinear circuit, such that the TWT's output signal seems undistorted. This conditioning relates to a controlled distortion of the signal according to the expected phase and amplitude distortion in the TWT. Thus, the circuit is configured to "fix" the nonlinearity of the AM-AM and AM-PM transfer characteristics, by offering, e.g., a gain expansion where the TWT exhibits gain compression and analogously correcting the phase shift. Integrating the gain expansion and compression into one transfer characteristic together with the phase correction, the combination of linearizer and TWT at best yields an ideal limiter. In reality, a non-ideal transition is found between the region of linear amplification and the upper limit. This is shown in Figure 2.6. Spectrally, the signal after predistortion shows the intermodulation products obtained by solely driving the TWT, but at a phase shift of 180° (at best) such that destructive interference occurs. There are many different types of both active and passive predistortion circuits and linearization methods, allowing a trade-off between performance, cost, and complexity. An extensive overview can be found in [19].

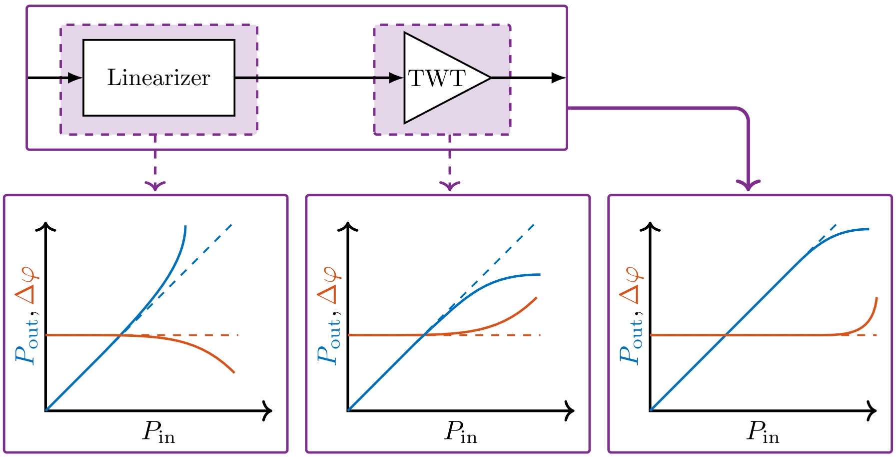


Figure 2.6: Schematic of a TWT with a preceding predistortion linearizer with individual and combined transfer characteristics.

Power Combining and Velocity Tapering

Another way to achieve a certain higher level of linear power is to increase the margin to saturation, for instance by increasing the available peak power in a controlled manner. A simple way to do so is utilizing power combining. While it significantly increases the power margin, using N TWTs does not necessarily result in an N-fold peak power. For efficient combination of several TWTs, the phase at the single TWT's output needs to be preferably stable over time and taken into account for the superposition of signals. Statistical variations in the output phase over drive, which are to be expected from, e.g., manufacturing tolerances, increase the power combining losses in addition to auxiliary elements required for the combining, such as couplers and further transmission lines that also introduce losses.

For the individual TWT, the peak power, linearity, and many other characteristics have in the past significantly been improved by means of velocity tapering [1], [20]. In modern helix TWTs the pitch profile of the delay line is inhomogeneously tapered to control the phase velocity of the wave and thus the synchronism with the beam. This can significantly increase the performance of the amplifier and strongly influence its behavior. The profiles nowadays usually stem from numerical optimization and thus, in principle, do not follow an analytic formulation.

The idea of tapering closely follows the operating principle of the TWT. As established, the average beam velocity decreases along the interaction region due to energy transfer towards the traveling wave. This relates to a change in synchronism and, thus, an altered beam-wave coupling, possibly impeding an efficient amplification

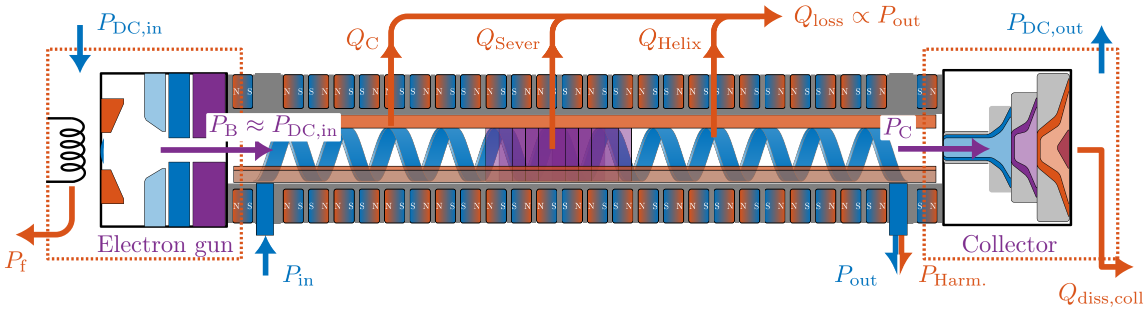


Figure 2.7: Schematic view of the power flow in a TWT.

process. Therefore, by velocity tapering, the circuit is varied along the beam axis, such that the synchronism is adjusted according to the decelerated beam. Early taper designs solely following this synchronism-adjustment approach focused on the area near the output to improve beam efficiency, leading to a higher peak power, but also to larger phase distortion. Later, more involved taper topologies have been utilized to also compensate the nonlinear phase shift [20]. This has lead to immensely improved TWT designs, both over drive and frequency, as well as against oscillations, and still serves as a fruitful source of TWT performance progression.

2.4 Efficiency and Power Flow

Typically, the power amplifier is one of the most power hungry components in a satellite. Therefore, its efficiency significantly contributes to the total efficiency of the system and is crucial for the overall power and thermal budget. Additionally, the efficiency plays a major role from an economic point of view, with a 1 % overall reduction leading to launch cost savings of more than 400 k€ per satellite[4]. Thus, it is important not only to identify the relevant loss contributors and mechanisms, but also to mitigate their impact.

In Figure 2.7, the power flow in a helix TWT is sketched. The collector and electron gun subsystems are each separated by a dotted line. All dissipated and harmonic powers are marked in red, the interfacing powers between the subsystems are shown in purple, and the RF- and DC-powers into and out of the TWT are marked in blue.

The heated cathode in the electron gun is powered by a heated filament. For this, a power P_{f} is required, which usually is in the order of a few watts. Also, a DC power supply is used to generate the current with an acceleration voltage V_0 relative to the

[4]This is assuming a typical satellite with 40 TWTs of 100 W output power each [4].

cathode. A beam power

$$P_\mathrm{B} = V_0 \cdot I_0 = P_\mathrm{DC,in} - P_\mathrm{f} \approx P_\mathrm{DC,in} \tag{2.18}$$

leaving the electron gun subsystem enters the interaction region.

Here, the beam interacts with the RF input power P_in, which is amplified along the delay line. In principle, the desired amplified signal and various harmonics arrive at the output coupler. This coupler typically is optimized for the amplification frequency band. Thus, large parts of the generated harmonics are reflected back into the device. While this plays a role for instability considerations, most of the time the reflected power will be absorbed in the sever without further consequences. The desired RF output power P_out is coupled out with a remaining harmonic part $P_\mathrm{Harm.}$. The amount of useful output power is related to the beam power by the beam efficiency

$$\eta_\mathrm{Beam} = \frac{P_\mathrm{out}}{P_\mathrm{B}}. \tag{2.19}$$

In addition, the total efficiency

$$\eta_0 = \frac{P_\mathrm{out}}{P_\mathrm{DC,in} - P_\mathrm{DC,out}} \tag{2.20}$$

covers the DC power regeneration in the collector, providing a more complete picture.

Along the delay line, various losses leading to an efficiency reduction are to be considered. First of all, a loss Q_Sever is produced in the severs. In addition, in the helical delay line, the barrel, and the support rods, material losses Q_C are generated. In parts, these are desired for stability reasons and depend on the RF power transported along the delay line. The reflected power at the output coupler also contributes to Q_C and Q_Sever[5]. Finally, especially for a strongly bunched beam, some electrons hit the delay line unchecked, locally generating dissipated power Q_Helix. Both Q_C and Q_Helix are unevenly distributed along the beam axis, being strongest at the end, as they depend on the beam modulation and local RF power. As the severs are usually located at positions far from the output, the total dissipated power in the attenuator is small compared to P_out, even though the total attenuation is at least in the order of tens of dBs. Therefore, while the total interaction region losses

$$Q_\mathrm{loss} = Q_\mathrm{C} + Q_\mathrm{Sever} + Q_\mathrm{Helix} \tag{2.21}$$

cannot exactly be calculated or accurately be predicted, in back-off they are roughly proportional to the output power and can be estimated to around 15 % of P_out [2].

[5] For the input coupler, this is negligible.

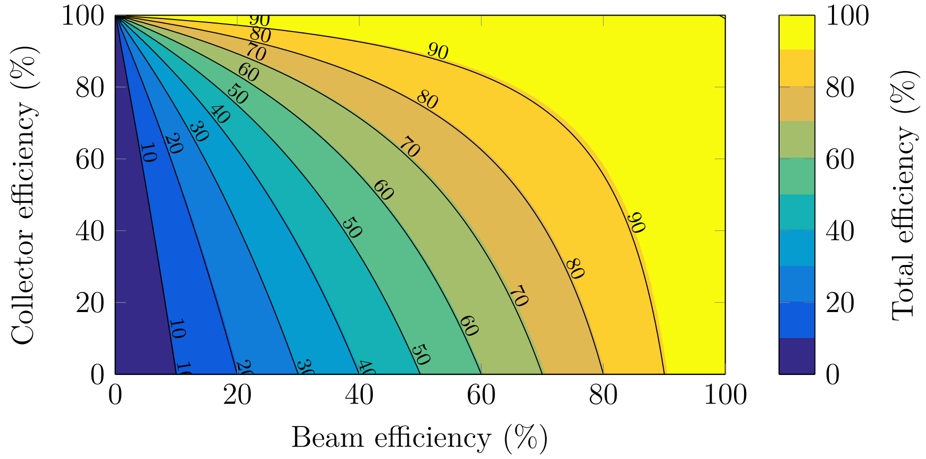


Figure 2.8: Approximate total efficiency of a TWT as a function of the collector and beam efficiency.

In modern designs, the collector losses

$$Q_{\mathrm{Coll}} = P_{\mathrm{B}} - Q_{\mathrm{loss}} - P_{\mathrm{out}} - P_{\mathrm{DC,out}} \tag{2.22}$$

reach similar dimensions. From this, a collector efficiency

$$\eta_{\mathrm{Coll}} = \frac{P_{\mathrm{DC,out}}}{P_{\mathrm{B}} - P_{\mathrm{out}} - Q_{\mathrm{loss}}} \tag{2.23}$$

can be defined, relating the DC output power reduction to the remaining power in the spent beam. With

$$\eta_0 \approx \frac{\eta_{\mathrm{Beam}}}{1 - \eta_{\mathrm{Coll}}(1 - \eta_{\mathrm{Beam}})} \tag{2.24}$$

a simplified relationship between collector and beam efficiency can be found, where filament and interaction region losses are neglected [3]. The impact of improving either η_{Beam} or η_{Coll} according to Equation (2.24) is shown in Figure 2.8.

3 Simulation Models for Modulated Signals

This chapter presents a selection of possible ways to simulate multi-tone and communication signals in of TWTs. In principle, there are various options to fulfill this task.

From the TWT manufacturer perspective, the simulation of well-established continuous-wave (CW) characteristics like single-frequency saturated input and output power or nonlinear phase shift have historically been most important for evaluating the amplifier performance. As established in the introduction, today, increasingly complex modulation schemes are used. Therefore, other performance metrics are on the rise, which also changes the requirements for simulation methods. In general, suitable methods for this task should be as fast as possible and as accurate as necessary. Therefore, for each of the models the accuracy of the results, the required computation time, and their practicability need to be considered.

Frequency-domain codes like MVTRAD [11] or CHRISTINE [12] are fast, reliable, and well-validated for CW simulation, but are naturally limited for more general time-varying signals. Still, in some cases, modified versions of such frequency-domain codes can be used for multi-tone calculations. Anyway, the more intuitive choice are time-domain based tools like CST Particle Studio [13] or DiMoHa [21], [22], which allow for arbitrary input signals, and thus also allow the study of stability issues, transients, or memory effects. For time-domain full-wave simulation tools, the biggest drawbacks however are the quite involved modeling required for accurate results and the very long computation times. Therefore, they are most suitable for validation and investigation of special phenomena. Alternative proprietary time-domain tools like DiMoHa, which are specialized on TWTs, are currently under development. While they are potentially faster than general purpose tools, they still have to come to age. Even then, a full simulation of, e.g., realistic QPSK modulated signals on multiple carriers would take quite a long time. Within the possible options for modulated signals, hybrid time- and frequency-domain envelope approaches are the middle ground, trying to combine the best of both worlds: a time-domain integration according to the modulated signal with a frequency-domain backbone.

In the following sections, a discussion of the several simulation methods regarding their performance, advantages, and disadvantages, as well as a comparison between each other and to measurement results is given. Also, the relation between some well-known CW quantities and multi-frequency quantities is briefly investigated.

3.1 Parametric Simulation with MVTRAD

MVTRAD, short for *MouVemenT RADial*, is a proprietary steady-state frequency-domain code. Following the so-called parametric approach, the slow-wave circuit is represented by cold parameters[1], such as the coupling impedance and the phase velocity. To extract these parameters, the eigenmodes of a periodic unit cell are calculated by full-wave simulation or analytic models, such as the tape-helix model [23], [24]. MVTRAD is specifically designed for the large-signal simulation of helix TWTs and, for this case, well-verified and established. In the following, its operation principle is briefly outlined, as it serves as a reference for many results and models in this chapter.

3.1.1 Operating Principle

The 2.5D PIC-code MVTRAD computes the beam-wave interaction by modeling the electron beam as macro-particles in a 2D plane and describing the RF helix fields by a number of harmonics. Rotational symmetry of both the electromagnetic fields and the beam is assumed, while still taking the azimuthal movement of the particles into account.

The circuit's eigenfields are built by means of field expansion. For this purpose, the harmonic expansion in Equation (2.3), and consequently the phase velocity from Equation (2.4) and the coupling impedance from Equation (2.7) are used. In fact, inside the beam region, in terms of beam-wave synchronism and field magnitude, mainly the space harmonic with $n = 0$ is relevant for the amplifier simulation. Thus, for a uniform helix geometry, the cold circuit properties can be represented by a single coupling impedance value and a single phase shift per frequency point. The RF fields are described by a number of harmonics according to the number of frequency components that are to be considered in the simulation. The cold parameters describing the whole

[1] "Cold" parameters are parameters of the structure without an electron beam. When an electron beam is considered, one speaks of, e.g., "hot" parameters.

slow-wave structure therefore need to cover all used frequencies and all geometric configurations of the delay line.

In addition to the circuit fields, magnetic focusing fields and the space-charge fields are taken into account. The latter are obtained numerically from the particle distribution. In this implementation of the PIC approach, these particles are pushed in time-domain, while the circuit fields are described in frequency-domain. In each time-step of the simulation, particles are emitted into the interaction region. Following the PIC-approach, their charges are distributed onto the numerical grid, representing the space-charge density inside the beam. In contrast to the circuit fields, the space-charge field is calculated at each time step, assuming it to be static during the time-step. This enables the use of Poisson's equation instead of the full description by Maxwell's equations, reducing the computational effort.

The particle trajectories resulting from the interaction with all considered fields are computed iteratively, leading to an update of space-charge and circuit fields. This is repeated until convergence is reached.

Any tool enabling meaningful simulation of modulated signals obviously needs to include all significant nonlinear effects as described in Section 2.3. In MVTRAD, the particle beam contains all necessary effects, as it is described as macro-particles and their motion is integrated in time-domain. As the number of particles required is significantly reduced by utilizing symmetries, for many cases MVTRAD is much faster than general purpose PIC tools. Typically, large-signal CW computations can be carried out in less than a minute per frequency and power setting on a typical desktop computer.

3.1.2 Multi-Tone Operation

Although any nonlinear mixing in the electron beam can in principle be simulated in MVTRAD, the electromagnetic fields are only considered in terms of known harmonics. While this indeed simplifies the field calculation for determining typical amplifier design characteristics from CW excitation, it complicates the multi-tone simulation and all the more the simulation of arbitrarily modulated complex signals. Therefore, the offered scope of applicability is limited. Still, MVTRAD can be used for multi-tone operation, as the general approach behind it allows for not only the fundamental frequency, but also harmonics of it. RF-harmonics are already required for CW excitation, as they are generated in the electron beam and contribute to the amplification process. This multi-harmonics capability can, in certain cases, be used for a multi-frequency excitation, where the input signal consists of several tones inside the amplification band. In principle, such a signal can be calculated by taking the greatest common divisor as the fundamental frequency with a sufficiently high number of included harmonics. This requires a possibly exorbitant number of harmonics to be considered, and as a result strongly increases the computation time. Also, the update frequency Δf_τ, as the inverse of the time step $\Delta\tau$ in which fields and particles are updated, is selected depending on the fundamental frequency $f_0 = \frac{1}{T_0}$ to

$$\Delta f_\tau = \frac{1}{\Delta\tau} = \frac{N_{\mathrm{div}}}{T_0} = f_0 N_{\mathrm{div}}, \tag{3.1}$$

with an integer scaling N_{div} between them. Thus, of course, the number of time steps per RF period for a multi-tone simulation needs further scaling to achieve a similar time resolution as in the single-tone computation case. As the greatest common divisor would generally be orders of magnitude smaller than the operating frequency - and for arbitrarily modulated signals in principle infinitely small - the calculation time would increase disproportionately. To counteract this, some adaptions have been included in MVTRAD. While the number of harmonics is immense, many of the harmonics are only dummies and not actually required for the field description. Therefore, these are ignored for the field updates. Also, while a high number of time steps per period is required, full particle injection into the calculation domain may possibly be required less frequently without significant influence on the overall particle density. Although these adaptions strongly reduce the computation time, the use of codes like MVTRAD for multi-tone simulation still is only possible in special cases, as is reported in Section 3.4.

3.2 Full-Wave Particle-In-Cell Simulation

Time-domain full-wave simulation tools allow arbitrary input signals and, thus, enable the study of a variety of time-domain phenomena. The biggest drawbacks of these geometry-driven solvers, however, are a significantly more involved modeling and long computation times. In this section, at first the operation principle is shortly introduced. Afterwards, some modeling approaches for one of the most prominent codes of this category, CST Particle Studio's PIC solver, are presented.

3.2.1 Operation Principle and Motivation

Full-wave PIC codes [2] follow the "first-principle"-approach. They directly solve Maxwell's equations in a geometry-driven strategy, fully in time-domain. Therefor, they combine many electrons to macro-particles and deposit the macro-particle charges on a grid to reduce the number of inter-particle force calculations. The coupled problem is then solved iteratively: From the particle velocities and positions, the particle charges are projected on the numerical grid. Using Maxwell's equations, the electric and magnetic fields on the grid are updated and weighted back to the particle positions. Then, using the Lorentz force law and the equation of motion, the particle movement is calculated and the particle positions and velocities are updated. This is repeated until the specified simulation time is reached.

CST Particle Studio's PIC solver is a well-known example for such a general purpose tool. It is regularly used in literature as a benchmark for new simulation tools or to predict the performance of novel TWT topologies, for instance planar delay-line topologies [25], [26] or FW-TWTs [27], [28]. In this work, CST's time-domain background is especially useful. It enables the study of transient processes or complex modulation signals, where frequency-domain codes like MVTRAD are naturally limited. While novel and promising time-domain codes based on similar parametric information of the slow-wave circuit are under development [21], [29], they are currently not yet fully validated, but similarly specialized as MVTRAD.

The usage of CST presents several challenges. Apart from the high demands on the computation side, the tube modeling for accurate results is fairly complex. Having a precise model and a reproducible modeling approach is a key requirement for using CST for the simulation of TWTs under modulated excitation. The following section considers some challenges and approaches and thus draws from the findings previously reported in [30], [31], where complete analyses of realistic TWTs with CST's PIC code have been carried out.

3.2.2 Modeling

In the following, models for two of the most challenging problems are outlined: the electron beam injection and the severs. The proposed approaches enable a quick and mostly automated model generation. Therefor, the same input files and model descriptions as for MVTRAD are used. To show the accuracy obtained with the shown approaches, CST results for commercial medium power Ku-Band helix TWTs for satellite downlink applications are compared to those of MVTRAD and to measurements.

Table 3.1: Currents from CST for simple geometries and realistic electron gun models, compared to suitable references.

	CST	Reference	Error (%)
Quadratic plates	65.9 A	65.4 A	0.76
(100 mm × 100 mm × 3 mm)	(36 min)	*Child-Langmuir*	
Concentric spheres	27.0 A	27.5 A	1.81
(Radii 40 mm, 37 mm)	(27 min)	*Langmuir-Blodgett*	
Concentric cylinders	0.595 mA	0.598 mA	0.50
(Radii 5 mm, 15 mm, length 10 mm)	(30 min)	*Langmuir-Blodgett*	
Electron gun model A	83.9 mA	81.3 mA	3.19
(realistic geometry)	(2:40 h)	*Gun2D (Thales)*	
Electron gun model B	74.1 mA	75.2 mA	1.46
(realistic geometry)	(2:45 h)	*Gun2D (Thales)*	

Electron Gun Simulation and Direct Beam Calculation

In contrast to a direct computation of the electron gun as a part of the PIC simulation, gun and delay line are generally considered separately. While this is far more efficient, it requires the injection of the particles into the interaction region from an artificial surface. As the operating principle of a TWT obviously relies on the electron beam, its proper generation is crucial. There are two practicable approaches for this task.

To simulate the electron gun itself, the three-dimensional structure of the electron gun is directly modeled in CST. Following a particle tracking simulation, this enables the investigation of various beam properties. The particle emission is based on the Child-Langmuir emission law, which is defined for the electron beam flow in parallel-plane diodes [1]. To evaluate the usability of CST for electron guns, both simplified structures and realistic gun geometries are investigated. For planar, parallel surfaces and arrangements of concentric spheres, analytic solutions are readily available from literature [1], [32]. In Table 3.1, these theoretical references are compared to the simulated results from CST, showing good agreement. For the comparison of the CST simulation of real gun geometries, the proprietary Thales code Gun2D is used as a benchmark. This reference tool is much faster, as it takes advantage of the rotational symmetry of electron guns and runs in 2D, which CST is not capable of. The total extracted beam current deviation is also reported in Table 3.1.

As an alternative, a direct injection of the beam according to specified beam properties can offer the desired beam with less effort. In that case, the electron beam is generated directly on an abstract particle emission surface. For rotating pencil beams, concentric emission rings are defined. On each of them, the particles are emitted at a given solid angle and with an initial energy. Alternatively, for full control over the injected beam, precedingly calculated emission points and initial velocities, e.g., using Equation (2.1), can be imported into CST. The injection points are distributed on the beam cross-section such that the charge density is constant. This can be seen in Figure 3.1a, where each injection point is also marked as the starting point of a corresponding velocity vector. The base color indicates the initial velocity.

Both with the direct electron gun simulation and the beam calculation, the beam data is to be recorded, e.g., in an ASCII-file. To include the recorded data into the PIC computation domain, an adequate interface between gun and interaction region is required. This is established in the following.

Particle Injection Interface

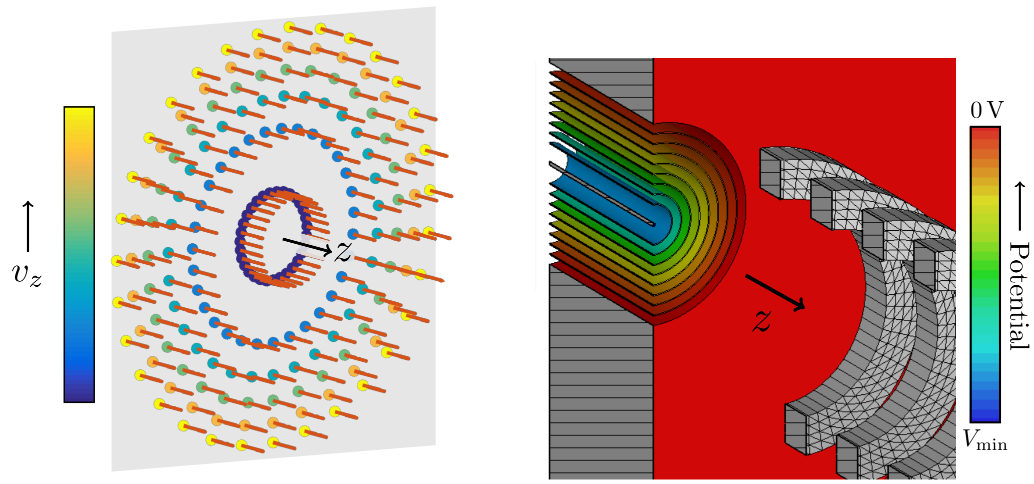
By default, the particle injection procedure in CST does not consider the static potential of the emitted beam as a necessary boundary condition at the injection plane. The macro-particles are simply pushed through an emission surface. Thus, the missing beam potential alters the outgoing beam, leading to a different beam shape and velocity distribution and ultimately a deviation in the trajectories.

To compensate for this, an additional boundary condition is included at the interface. It stems from the assumption of an infinite beam extension in axial direction inside a beam tunnel. Near the injection plane, the electron beam is not yet modulated. Thus, the helix can be approximated by a hollow, perfectly conducting cylinder. Therefore, a rotationally symmetric potential distribution

$$V(r) = \frac{\rho_0 r_\mathrm{b}^2}{2\epsilon_0} \begin{cases} \left(\frac{1}{2} + \ln\frac{r_\mathrm{h}}{r_\mathrm{b}} - \frac{r^2}{2r_\mathrm{b}^2}\right), & r < r_b, \\ \ln\frac{r_\mathrm{h}}{r}, & r \geq r_b, \end{cases} \tag{3.2}$$

is defined in the transverse emission plane, resulting from direct integration of Poisson's equation [33], where r_b denotes the beam radius and r_h the helix radius.

The resulting electrostatic field is generated in a particle-free electrostatic simulation, with a configuration of concentric rings in the emission plane as shown in Figure 3.1b. Each ring is fixed at a specified potential, such that the equivalent boundary condition



(a) Distribution of injection points in the injection plane with corresponding initial energy vectors. The color denotes the initial velocity v_z in z-direction.

(b) Setup for the electrostatic simulation with a defined number of potential rings in the injection plane. The outer hull at zero potential is not shown.

Figure 3.1: Injection of particles and equivalent beam port with potential rings.

from Equation (3.2) is fulfilled. A PIC simulation without including this external field from the electrostatic simulation would introduce a significant voltage depression in axial direction. This causes an error in the synchronism between the traveling wave on the delay line and the electron beam, which directly leads to an error in the predicted behavior of the TWT. The resulting extended workflow representing the proposed beam port is summarized in Figure 3.2.

Sever Modeling

Typical severs are, for instance, realized by nanometer-scale material layers applied to the support rods. Realistic, physical modeling of such a structure is thus not feasible in CST. Alternatively, equivalent lossy materials are defined to emulate the desired loss properties. Therefor, in this work, a nonphysical "lossy vacuum" is placed outside of the helix as the contributor of the virtual equivalent losses. In principle, the helical wire or the support rods could also be chosen for this purpose. The lossy-vacuum material is defined with a constant relative permittivity $\epsilon_r = 1$ to ensure minimal reflection, but a loss tangent $\tan(\delta)$ tapered in axial direction. This taper is implemented with a large number of slices, each with constant $\tan(\delta)$. This is sketched in Figure 3.3.

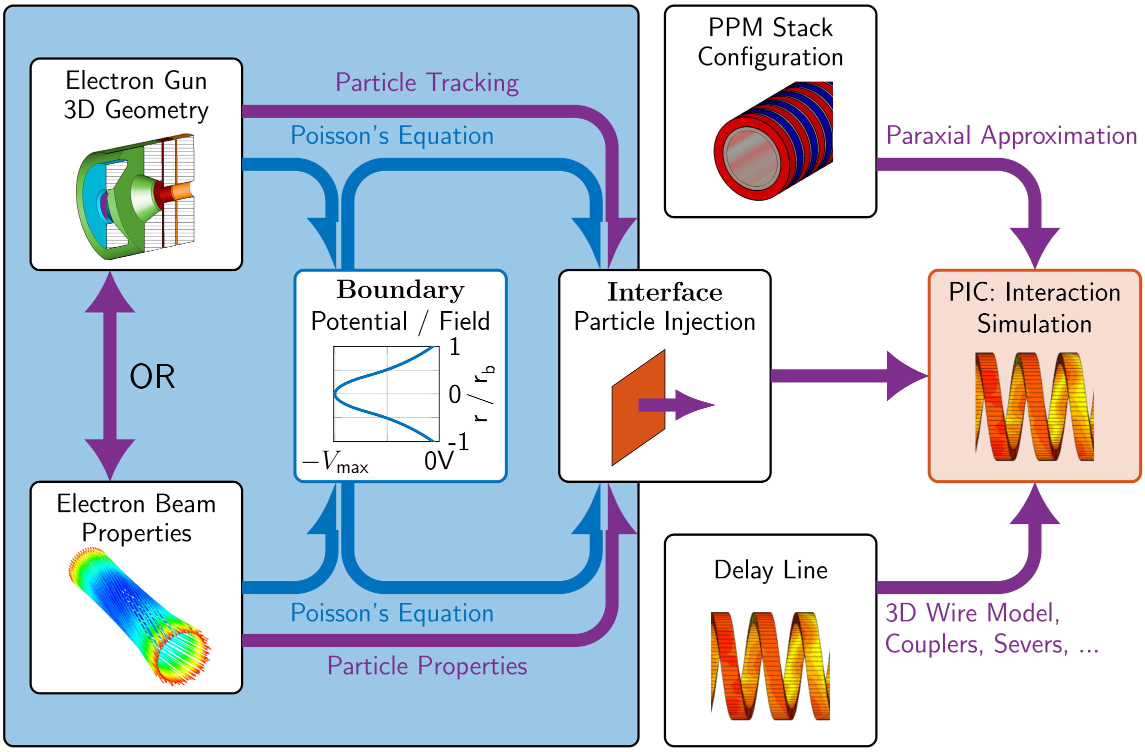


Figure 3.2: Workflow for the beam port inside the PIC interaction simulation in CST Particle Studio.

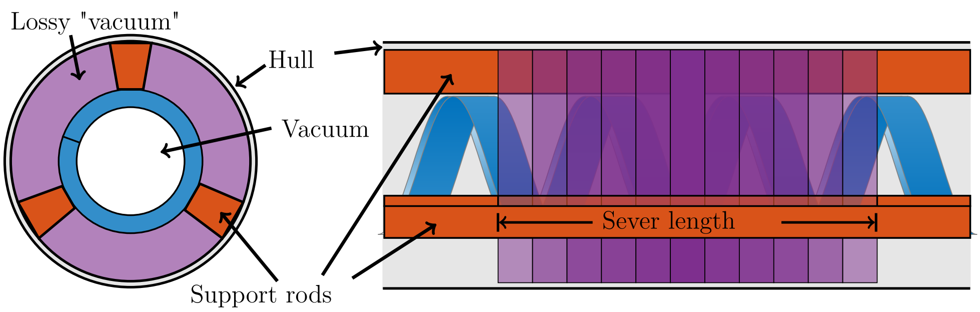


Figure 3.3: Equivalent loss mechanism model. The transparency of the sever slices represents the loss profile. The losses per slice decrease symmetrically from the center of the sever towards the ends.

The loss profile $\alpha(z)$ of the sever, e.g., from measurements, is translated into a material distribution

$$\epsilon(z) = \epsilon_0 \epsilon_r \left[1 - j \tan(\delta(z))\right], \tag{3.3}$$

such that the absorption is substituted accurately. For this purpose, a perturbational approach is adopted to find a relation between loss and material profile. Therefor, several assumptions are made. Most fundamentally, one assumes that the introduction of losses does not relevantly alter the field distribution. This allows the initial calculation of field properties without considering lossy materials, and significantly reduces the computational effort. As a consequence, the result's validity is limited to a region around the initial field calculation at a reference value $\tan(\delta_0)$. The extent of this region depends on the considered scenario, the degree of simplification in the approach and on the required accuracy of the translation.

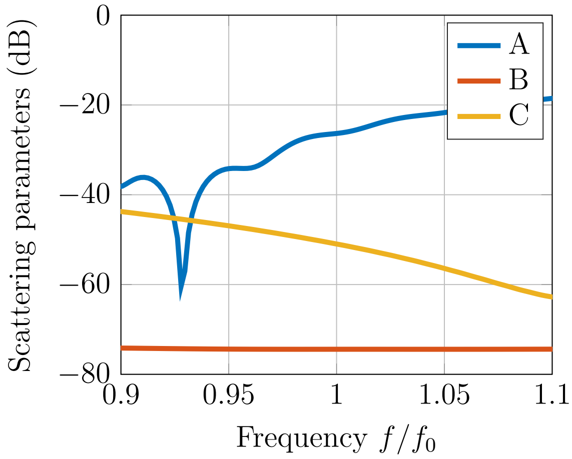
A similar fundamental approach can be found in literature [34], where the conductivity losses of the helix wire instead of the losses in the sever are considered. It assumes that introducing the lossy material does not affect the field distribution itself and thus can be based on results of an unperturbed eigenmode simulation. In case of [34], this assumption obviously holds, as the losses in general are small. Here, it might be violated in the high-loss center of the sever. Fortunately, experience with MVTRAD shows, that such inaccuracies are negligible compared to the accuracy at the low-loss outskirts of the sever. For well-matched severs, very smooth transitions are important. Thus, the transition from loss-less to highly lossy is rather long. Along this transition, the fields do not change strongly.

In addition, at the sever positions, the beam is not yet strongly modulated, meaning that the fill-factor is not much different from the one at the injection. Like in MVTRAD, one can assume, that it suffices to consider the field properties on the beam axis. This, of course, also holds for the lossy regions.

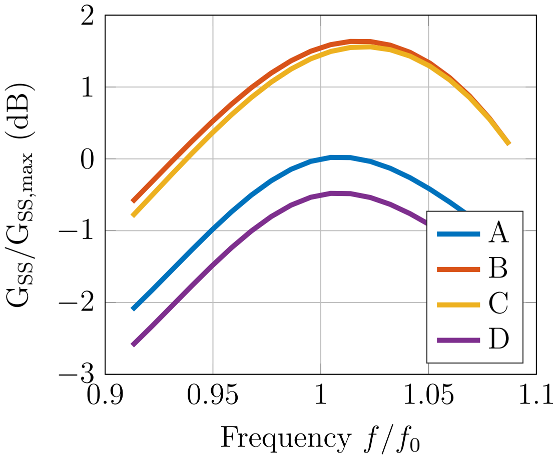
For the slices, the surface and volume losses, which can be calculated separately, can be summarized by the total dissipated power P_L. Surface losses, e.g., due to the wire material, are assumed to be invariant in axial direction and small in comparison to the volume losses, which stem from the complex permittivity $\epsilon(z)$. By means of P_L, the total stored energy W can be related to the quality factor Q, which describes the energy dissipated during one period t_0, through

$$Q(\delta) = \frac{2\pi W}{P_\mathrm{L}(\delta) t_0} \Rightarrow W = W_0 \exp\left(-\frac{2\pi f}{Q(\delta)} t\right), \tag{3.4}$$

weighted by an arbitrary initial energy scale W_0. In a homogeneous slow-wave structure segment, the differential energy change along the axis can be related to the energy



(a) Performance of a sever with 75 dB total attenuation from the perturbational approach. (A): reflection S_{11}, (B): transmission S_{21}, (C): time-gated reflection S_{11}.



(b) Influence of the sever onset on the small-signal gain. (A): Reference TWT, (B): steeper transition of the loss taper, (C): smaller total attenuation, (D): larger total attenuation.

Figure 3.4: Sever performance and sensitivity to sever profile and total attenuation.

change over time and the energy transport velocity, which is equivalent to the group velocity v_g.

Assuming an exponential decay of the fields along the axis with $\exp(-\alpha z)$ we can define the attenuation constant α. Hence, with $P \propto |E|^2$, the power decays according to $\exp(-2\alpha z)$. Together with Equation (3.4) and a group velocity based substitution from time derivatives to spatial derivatives, a translation between material and loss profile is

$$\alpha(z) = \frac{P_\mathrm{L}(\delta)\big|_{\delta=0}}{2Wv_\mathrm{g}} + \frac{P_\mathrm{L}(\delta)\big|_{\delta=\delta_0} - P_\mathrm{L}(\delta)\big|_{\delta=0}}{2Wv_\mathrm{g}\tan(\delta_0)}\tan(\delta(z)), \tag{3.5}$$

which can be applied in an eigenmode simulation as part of a pre-processor routine.

This expression is rather simple, reducing the possibly complex relationship between α and ϵ to a very manageable one. This is a direct consequence of the manifold assumptions made. As discussed before, especially for the high-loss center, this might at first seem worrisome, and thus requires thorough consideration. A particle-free simulation reported in Figure 3.4a shows deviations of the total attenuation of around 3 dB for a sever with 75 dB loss, indicating that Equation (3.5) is still well suited. It leads to a smooth transition between interaction region and sever, which yields input reflections of less than −40 dB throughout the band.

Although the total attenuation is represented more or less correctly following above procedure, in the sever it is not its most important property. In comparison, Figure 3.4b shows that the profile at the sever onset is far more critical, as it effectively changes the electric lengths of the sections. Here, the small-signal gain of a reference TWT including two severs (curve A) is compared to that of the same TWT with a different loss profile on the second sever. This other loss profile relates to a typical simplification of the shape as utilized in proprietary tools. The small-signal gain is normalized to the maximum small-signal gain of the reference TWT. A steeper transition (curve B), and thus a different profile at the onset, can easily result in an error as large as if the total attenuation were chosen 50 dB smaller (curve C). This is mainly due to the changed effective electric length of each section. A 50 dB larger total attenuation (curve D) does not have a significant influence on the gain. As the losses are small close to the onset, the perturbation approach is accurate enough there and good agreement in terms of dissipated power can be expected.

3.2.3 Model Validation and Simulation Accuracy

To validate the presented modeling approaches, a variety of simulation runs and comparisons have been carried out. In the following, simulation results for a Ku-Band helix TWT are shown. A comparison of CST simulation with measurement results and the proprietary code MVTRAD is carried out. The CST model is set up in such a way, that it uses exactly the same input files describing setup and geometry as MVTRAD. The 150 W Ku-Band TWT considered here has a tapered, multi-section helical delay-line. Figure 3.5 shows the interaction region model as generated and simulated in CST.

For validation against measurements, the measured TWT needs to be known in as much detail as possible. In MVTRAD the phase velocity and the coupling impedance can be fitted to measurements directly. Fast run times enable the identification of unknown parameters by, for instance, parametric sweeps. In contrast, a model fitting in CST requires a physical reasoning. Thus, geometric measurements, such as the wire cross-section and shape, and also the contact between wire and support rods, are used in the modeling procedure to minimize deviations.

By this, mechanical deformations and similar imperfections, which significantly influence the phase velocity among other properties, are taken into account. While the sever profile near the onset is measured and translated into a material profile to be used in CST, the total attenuation is only an estimate, as the measurement of its exact level in the high-loss center is quite involved.

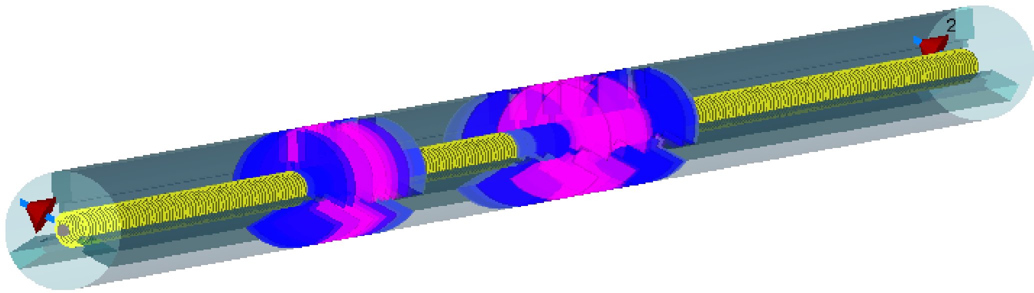
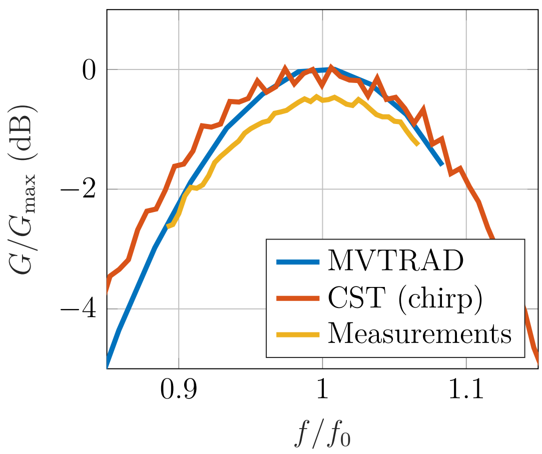


Figure 3.5: Full model of the interaction region with the tapered helix wire, two severs, support rods, and discrete ports at in- and output.

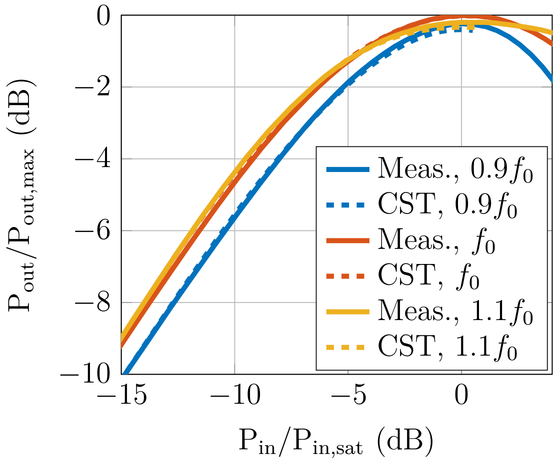
Information on the measured magnetic focusing field is imported on the beam axis and the magnetic field at the cathode is estimated from simulation. A large-signal convergence study returns around 200 injection points for accurate discretization of the electron beam. Also, the support rod permittivity is fine-tuned to adjust for other uncertainties such as material tolerances and to correct the phase velocity, resulting in a 2 % permittivity deviation from the nominal value. This is well within typical material deviations for the support rods [35].

Finally, precise full-wave simulation requires a high mesh density. It is initially investigated in particle-free simulation as an involved convergence study including the electron beam is not feasible due to long computation times. The convergence is assessed with respect to the electric field on the axis. Discrete ports are chosen to couple energy into or out of the interaction region. They correspond to simple and well localized current sources between the helix wire and the outer hull of the TWT. In contrast, a realistic and well matched coupler geometry requires long structures perpendicular to the axis. This results in a large number of additional cells in the structured mesh over the whole length and therefore lowers the number of emission points that can be used in the PIC simulation, due to the computation resource competition of the particle and mesh-cell number.

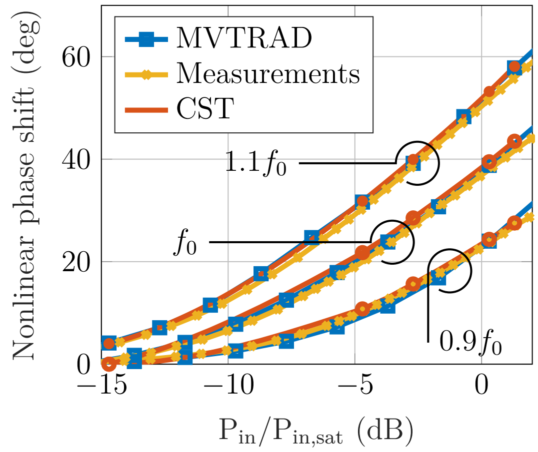
Figure 3.6 shows a number of simulated properties from CST and MVTRAD and the respective measurements. MVTRAD assumes perfectly matched couplers at input and output. For Figure 3.6a, a chirp signal is used to reduce the computation time from roughly half a day per frequency point to less than a day for the whole frequency band. The gain is slightly overestimated by around 0.5 dB both in MVTRAD and CST, which can be explained from the generally high sensitivity of the gain as outlined above. A corrective shift of the gain curve shows that MVTRAD slightly underestimates the



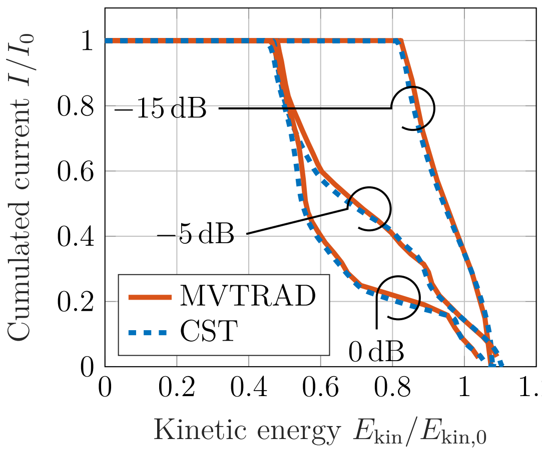
(a) Small-signal gain versus frequency.



(b) Output power versus input back-off.



(c) Nonlinear phase versus input back-off.

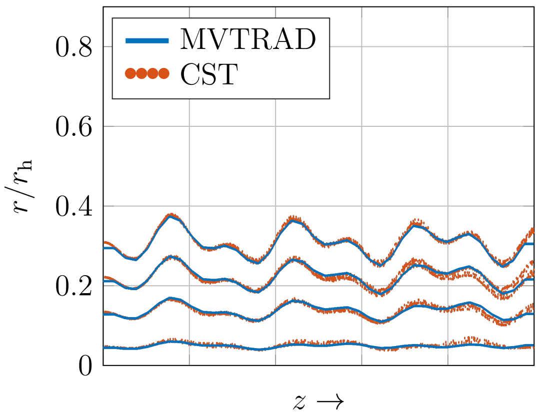


(d) EVS at three different back-off levels.

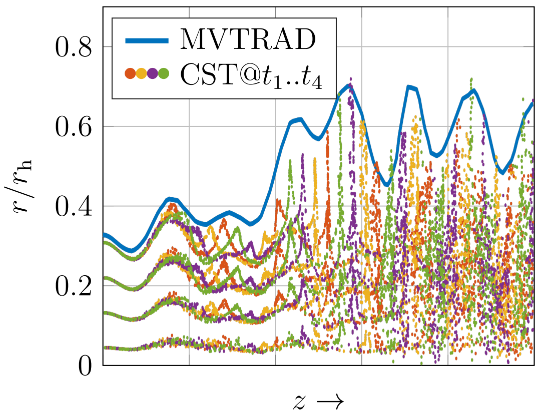
Figure 3.6: Comparison of different properties of the TWT from simulation in CST and MVTRAD with measured results.

bandwidth, especially visible at the lower band edge. This is not the case for CST, where the deviation is approximately constant.

As the small-signal gain computation exhibits a slight offset, the gain compression and the nonlinear phase shift shown in Figure 3.6b and Figure 3.6c are normalized to the respective saturated input powers. They are plotted versus input back-off at the band edges and at the center frequency. The error in saturated output power is within 0.15 dB at the edges and significantly smaller in-between. Similar agreement can also be seen for the nonlinear phase shift. The simulated values from MVTRAD and CST are close to the measured ones, with a maximum error of below 3°.



(a) Zero-drive excitation.



(b) Saturation.

Figure 3.7: Comparison of macro-particle trajectories at f_0 from CST and MVTRAD.

Figure 3.6d compares the electron energy spectrum at the output of the interaction region. Here, the accumulated current is plotted against the kinetic energy of the macro-particles for small-signal excitation, for saturation, and at an intermediate power level. Especially at high energies, the results from CST and MVTRAD are close. The energy of the slowest particles, which typically determines the lowest collector stage potential, is around 1 % to 3 % smaller in CST than in MVTRAD. As the electron beam spectrum is not easy to measure directly, it is not clear whether the CST or MVTRAD results are closer to reality. This is also the case for the particle trajectories shown in Figure 3.7, which visualize the particle motion inside the interaction region with a PPM-stack for focusing. For this purpose, the number of injection radii is reduced to four. Figure 3.7a and Figure 3.7b show the trajectories at zero-drive and at saturation, respectively. The trajectories are almost identical at zero-drive. For the high-drive case, several snapshots of the particle distribution are taken within one RF-period. They are compared to the beam envelope extracted from MVTRAD. Except for some peaks visible from the CST results, the general shape of the beam is represented accurately.

Overall, the CW interaction simulation results, i.e., gain compression, nonlinear phase shift, and the EVS agree well with measurements and an MVTRAD model with directly fitted parameters. The shown efficient approaches for the different parts of a full-wave helix TWT simulation in CST lead to models which are usable for the study of multi-frequency simulation in later sections of this work, especially Section 3.4.

3.3 Envelope Models

In this work, methods, in which any output quantity is calculated based on the signal envelope or related input signal characteristics, are called Envelope methods. TWT models in which Envelope methods are utilized to estimate the amplifiers response are so-called Envelope models, which enable the combination of frequency-domain and time-domain approaches to simulate the behavior of the TWT when excited with arbitrary time signals. The general idea behind Envelope methods is that for any small time interval, the TWT can be considered operating in a short-term steady state. Thus, for each of these intervals, one steady-state response can be found. This steady-state response can, for instance, be extracted from a frequency-domain calculation or a CW measurements. This way, the full signal can be integrated fast in time-domain.

In simplified terms, the envelope of a signal is a curve following the outlines of the signal's oscillation through its extremes. Regarding a modulated carrier

$$x(t) = A(t)\cos\left(\omega_0 t + \varphi(t) + \varphi_0\right), \tag{3.6}$$

with the carrier frequency ω_0, the non-negative modulated amplitude $A(t)$, the phase modulation $\varphi(t)$, and an initial phase shift φ_0, the complex representation or complex envelope form [36]

$$x(t) = \mathrm{Re}\left\{A(t)e^{j\varphi(t)}e^{j\omega_0 t+\varphi_0}\right\} \tag{3.7}$$

describes the signal by separating carrier and modulation[2]. Following this representation, the complex carrier $e^{j\omega_0 t+\varphi_0}$ is modulated in time according to the complex envelope function $A(t)e^{j\varphi(t)}$.

In simple Envelope method based approaches, a complex-valued transfer function describes amplitude compression and phase distortions caused by the nonlinear device. A set of amplifier transfer characteristics between input and output of the TWT, such as the AM-AM and AM-PM curves, can most conveniently be used as a reference for the model synthesis. This is well established and has led to various simple and fast approximate TWT descriptions, which work reliably under certain constraints. The most crucial requirement is that the time-domain input signal varies significantly slower than any transient effects and time constants in the TWT. Otherwise, one could not assume a steady-state response along the signal envelope. The TWT's time constants mainly stem from its electrical length and the retroactive length of the electron beam. Thus, they are at most in the order of a few nanoseconds. Furthermore, the dispersive properties of the TWT need to be negligible within the considered frequency range of

[2]Although $x(t)$ is still not a complex signal this is the typical terminology found in literature.

the signal. This is especially true for more simple descriptions of the Envelope method, as these do not include any frequency-dependent properties. As helix TWTs developed for communication systems are only marginally dispersive in practical frequency bands, this complies in many cases.

In the following, an overview over different frequency-dependent and -independent methods is presented, together with selected implementation details and comparisons.

3.3.1 Analytic Methods and Simple Two-Tone Estimates

Classically, the multi-tone performance of a TWT is described in terms of typical two-tone characteristics. The spectral components at the excitation frequencies f_1 and f_2 in the output signal relate to those that arise in the TWT with $nf_1 + mf_2 > 0$, for $(n, m) \in \mathbb{Z}^2 \setminus \{(0, \cdot), (\cdot, 0)\}$. These components are called intermodulation products and are a result of the nonlinearity as described in Section 2.3. They can be interpreted as distortions of the excitation and are desired to be small for TWTs used for instance in satellite communications.

One well-established and simple characteristic is the intermodulation product ratio D^i_{2n+1}. It can be calculated from

$$D^{\{1,2\}}_{2n+1} = 10 \log_{10} \left(\frac{P_2 \left(f_{1,2}\right)}{P_2 \left(f_{1,2} + (-1)^{\{1,2\}} n \, \Delta f\right)} \right), \tag{3.8}$$

where $\Delta f = f_2 - f_1 > 0$ is the spacing between the inserted tones (and therefore determines the intermodulation ratio) and $P_2(f_i)$ is the output power at the frequency f_i. In this section, only narrowband scenarios are considered, such that the TWT's dispersion can be neglected. Thus, with symmetric excitation, $D^1_{2n+1} = D^2_{2n+1}$ holds, as the spectrum at the output then also is fully symmetric.

There are several ways to estimate the intermodulation product ratio from the CW characteristics of the TWT. In the following, an operator M describing the TWT is considered, such that the output signal is given by

$$y(t) = (M \circ x)(t). \tag{3.9}$$

For a two-tone input signal

$$x_{\omega_0, \omega_0 + \Delta\omega}(t) = A \cos(\omega_0 t) + A \cos((\omega_0 + \Delta\omega)t + \varphi_0) \tag{3.10}$$

with equal amplitudes A, an arbitrary initial phase φ_0, and a frequency spacing of $\Delta\omega \ll \omega_0$, in principle a common period T is not always available. Still, for the non-dispersive case, one can assume that one can equivalently consider a signal $x_{\hat{\omega}_0, \hat{\omega}_0 + \Delta\omega}$

with $\hat{\omega}_0 \approx \omega_0$ and $\hat{\omega}_0 = p\Delta\omega, p \in \mathbb{N}$. By this, the signal $x_{\hat{\omega}_0,\hat{\omega}_0+\Delta\omega}$ is periodic with

$$T = \frac{2\pi p}{\hat{\omega}_0} \tag{3.11}$$

and

$$P_{2,\omega_0}(\omega_0 + n\Delta\omega) \approx P_{2,\hat{\omega}_0}(\hat{\omega}_0 + n\Delta\omega) \tag{3.12}$$

holds, where $P_{2,\omega_i}(\omega_j)$ is the output power at ω_j after M, when the excitation is $x_{\omega_i,\omega_i+\Delta\omega}$ and $n \in [-p; \infty[$. Using the signal's periodicity, the output signal can be found by a Fourier series expansion

$$y_{\omega_i,\omega_i+\Delta\omega}(t) = \left(M \circ x_{\omega_i,\omega_i+\Delta\omega}\right)(t) \tag{3.13}$$

$$\stackrel{!}{=} \sum_{n=1}^{\infty} \left[a_n \cos\left(n\Delta\omega t\right) + b_n \sin\left(n\Delta\omega t\right)\right], \tag{3.14}$$

with coefficients

$$a_n = \frac{2}{T} \int_0^T \left(M \circ x_{\omega_i,\omega_i+\Delta\omega}\right) \cos\left(n\Delta\omega t\right) \mathrm{d}t, \tag{3.15}$$

$$b_n = \frac{2}{T} \int_0^T \left(M \circ x_{\omega_i,\omega_i+\Delta\omega}\right) \sin\left(n\Delta\omega t\right) \mathrm{d}t. \tag{3.16}$$

The respective intermodulation ratio results in

$$D_{2n+1}^{(i+1)} = 10 \log_{10} \left(\frac{a_{p+i}^2 + b_{p+i}^2}{a_{p+i+(-1)^{i+1}n}^2 + b_{p+i+(-1)^{i+1}n}^2} \right), \tag{3.17}$$

with $i = 0$ for the lower and $i = 1$ for the upper intermodulation ratio.

While this description in principle also holds for any non-static operator, considering a purely static gain enables describing M by simple series expansions. In such a static gain description, the output signal $y(t)$ instantly follows the input signal $x(t)$. Consequently, there is no phase at all, as phase can only be defined by frequency or time evolution. This is a rather strong constraint, but it allows, for instance, to extend M by terms of a Taylor series. That would lead to expressions for the intermodulation products solely based on the parameters used to fit the Taylor series to the static gain function and the amplitude of the input tones.

In [37], a non-static description is chosen for the two-tone estimation, expanding around the operating point of the applied signal. In the case of small signal amplitudes compared to saturation, this enables the approximate calculation of certain TWT output quantities. Also, phase information is included in the model.

With a differential gain compression

$$c(f) = 1 - \frac{P_{\mathrm{in}}(f)}{P_{\mathrm{out}}(f)} \frac{\mathrm{d}P_{\mathrm{out}}(f)}{\mathrm{d}P_{\mathrm{in}}(f)}, \tag{3.18}$$

and the differential AM-PM conversion coefficient

$$k_{\mathrm{p}}(f) = P_{\mathrm{in}}(f) \frac{\mathrm{d}\Theta(f)}{\mathrm{d}P_{\mathrm{in}}(f)}, \tag{3.19}$$

with the nonlinear phase shift Θ, the third order intermodulation ratio for two equally strong tones can be approximated by [37]

$$D_3 \approx -10 \log_{10} \left(\frac{c^2}{4} + k_{\mathrm{p}}^2 \right). \tag{3.20}$$

Another investigated quantity is the so-called phase-transfer or AM-PM transfer factor k_{T}. It characterizes the effect of changing the input power level of a strong signal at one frequency on the phase of a second, weaker carrier at another frequency according to

$$k_{\mathrm{T}}(P_{\mathrm{in}}) = P_{\mathrm{in}}(f_1) \frac{\mathrm{d}\Theta(f_2)}{\mathrm{d}P_{\mathrm{in}}(f_1)}, \quad P_{\mathrm{in}}(f_1) \gg P_{\mathrm{in}}(f_2). \tag{3.21}$$

From [37], an approximation for k_{T} can be found with

$$k_{\mathrm{T}}(P_{\mathrm{in}}) \approx k_{\mathrm{p}} + P_{\mathrm{in}} \frac{\mathrm{d}}{\mathrm{d}P_{\mathrm{in}}} \left[\arcsin \left(\frac{k_{\mathrm{p}}}{\sqrt{\left(1 - \frac{c}{2}\right)^2 + k_{\mathrm{p}}^2}} \right) \right]. \tag{3.22}$$

3.3.2 Static-Gain Curve Model

In the past, straightforward two-tone approximations as given in Section 3.3.1 have shown to be reliable compared to measurements [37], [38]. They can be used to effortlessly characterize the degree of nonlinearity in limited frequency ranges, based on typical quantities which are anyway considered in the design and measurement of a TWT. Unfortunately, it is not clear how meaningful simple two-tone characteristics really are for real modulated signals in narrow or broad frequency bands. Therefore, a model enabling the direct integration of an arbitrary signal is desired.

To derive the two-tone expressions, the TWT is assumed to be frequency-independent within a selected spectral range. Thus, a simple non-dispersive nonlinearity is considered for the amplifier model, which in principle can be used in a more generalized setting with any excitation in time-domain, especially complexly modulated communication signals. In literature there are several approaches using the frequency-independent, nonlinear gain compression and phase shift to estimate the intermodulation products for multi-tone signals, e.g., the IMAL code [39]. In this thesis, the approach is used in a slightly modified way to generate the output signal as well as specific TWT properties like the EVS and various communication characteristics.

Model Realization

In the following, an implementation of this method is described, which also serves as a base for the nonlinear description of the frequency-dependent models in the later sections. Thus, the following description is also relevant for the methods in Section 3.3.3, Section 3.3.4, and Section 3.3.5.

The goal is to include the developed simulation codes during the design process, for instance for fast evaluation of, e.g., a helix pitch profile in a delay-line taper optimization routine. Therefore, the model identification scheme needs to be highly efficient. The nonlinear and frequency-dependent performance is obtained from steady-state simulation codes such as MVTRAD, where the reference AM-AM and AM-PM curves are generated within minutes. The CW gain and phase are calculated in input-power intervals of ΔP_{in} and interpolated when needed.

In this method, the input signal $x(t)$ is assumed to be a single-tone oscillation, the amplitude of which changes slowly over time, in accordance with the constraints of the Envelope method. This time-dependent amplitude can be extracted from the signal itself by means of the Hilbert transform [40], [41]

$$\mathbb{H}\{x\}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} \mathrm{d}\tau = x(t) * \frac{1}{\pi t}, \tag{3.23}$$

which is used to construct the so-called analytic signal

$$x^{\mathrm{inst}}(t) = \sqrt{x^2(t) + \mathbb{H}\{x\}^2(t)} \cdot \exp\left(j \arctan\left(\frac{\mathbb{H}\{x\}(t)}{x(t)} \right) \right). \tag{3.24}$$

The Fourier transform of this description is purely single-sided. Thus, the implementation can be carried out by discrete-time Fourier transform, where the negative frequency components are suppressed [42]. The analytic signal can be split into an instantaneous phase term

$$\xi^{\mathrm{inst}}(t) = \arg\left\{ x^{\mathrm{inst}}(t) \right\} \tag{3.25}$$

and an instantaneous amplitude

$$A^{\mathrm{inst}}(t) = \left| x^{\mathrm{inst}}(t) \right|, \tag{3.26}$$

which corresponds to the envelope of the original signal $x(t)$.

The output signal $y(t)$ can then be calculated by this static-gain curve (SGC) model to

$$y(t) = g_{\mathrm{AM}}\left(\left|x^{\mathrm{inst}}(t)\right|\right) \cdot \exp\left(j\xi^{\mathrm{inst}}(t) - \Theta_{\mathrm{PM}}\left(\left|x^{\mathrm{inst}}(t)\right|\right)\right), \tag{3.27}$$

where $g_{\mathrm{AM}}(A) \neq \mathrm{fct.}(\omega)$ describes the input-amplitude dependent output-amplitude characteristics and $\Theta_{\mathrm{PM}}(A) \neq \mathrm{fct.}(\omega)$ the nonlinear phase shift of the amplifier.

Terminology-wise, naming this approach a static-gain curve model is not entirely correct. In contrast to the static-gain description in Section 3.3.1, the usage of a Hilbert transform and analytic signal description leads to an inherently memory-based model, where the linear and nonlinear phase information is taken into account. Still, for the sake of consistency with [43], [44], the name is maintained for this work.

Frequency Reference

The general idea of the SGC model is to assume non-dispersive behavior over a certain time-interval. Thus, a single-frequency model can determine the result according to Equation (3.27). A question that remains is the choice of the frequency which is best for the considered interval. In principle, a variety of choices is conceivable for the reference frequency. One obvious solution is the center frequency

$$f^{\mathrm{ctr}} = \frac{f_{\max} + f_{\min}}{2} \tag{3.28}$$

of the investigated signal, where $f_{\max}$ and $f_{\min}$ are the upper and the lower boundary of the power-carrying frequency range. In a complexly modulated signal, this would also be the carrier frequency, although there might be better choices, depending on where the carrier is located in the amplification band. Alternatively, the reference frequency could be some kind of dominant frequency, which does not necessarily need to be in the center of the band. This could be useful when, e.g., two tones are considered with strongly asymmetric power distribution, or when one tone is at a frequency where significantly less gain is to be expected.

Equation (3.25) introduces the instantaneous phase. From its derivative, a quasi-instantaneous frequency f^{inst}

$$f^{\mathrm{inst}} = \frac{1}{2\pi}\frac{\mathrm{d}\xi^{\mathrm{inst}}(t)}{\mathrm{d}t} \tag{3.29}$$

can be calculated, which represents a changing effective frequency over time.

A comparison of using this more involved frequency-identification process and the simpler dominant-frequency reference has been carried out. Figure 3.8 shows exemplary results. The used reference Ku-Band TWT was designed to be more

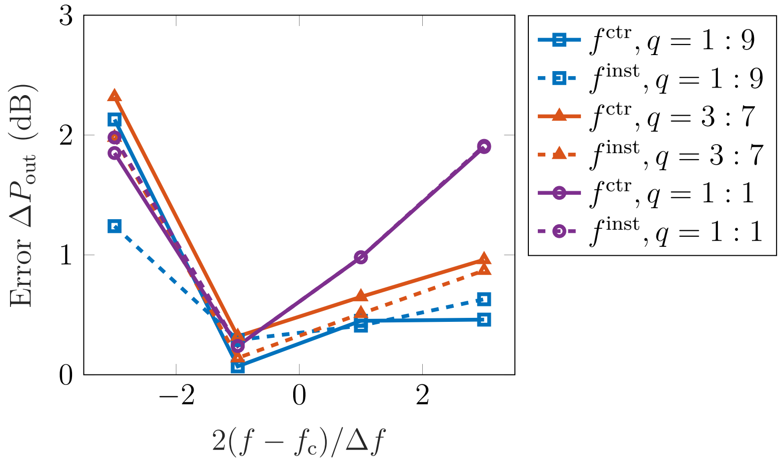


Figure 3.8: Comparison between the two reference frequency decision schemes for a large-signal simulation at $\Delta f = 500\,\text{MHz}$ around the amplification band center f_c. The error is compared to reference simulation carried out in MVTRAD.

dispersive than a typical communication TWTs, to highlight the effects. The figure shows three different constellations for two tones, which are $\Delta f = 500\,\text{MHz}$ apart, centered around the amplification band center at f_c. Apart from a symmetric excitation, two asymmetric cases are presented, with q describing the ratio of the powers at the two input frequencies. The simulation is carried out at a large-signal operating point near saturation of the band center. Thus, the presented scenarios are challenging for any Envelope model, as they cover large frequency ranges at a critical power level.

For asymmetric excitation the usage of f^{inst} shows advantages compared to the center frequency. For $q = 1:9$, the picture is not completely clear, as the results for the weaker main tone and the upper intermodulation product are slightly worsened by the instantaneous frequency choice, but the lower intermodulation product, which was far off before, is significantly improved. An improvement over the whole band can be seen for $q = 3:7$, where for all considered frequencies the instantaneous frequency selection shows a $0.1 - 0.3\,\text{dB}$ lower error. However, for the symmetric excitation, no significant difference can be identified. This can be explained by the progression of the instantaneous frequency itself. For a two-tone signal with amplitudes A_1 and A_2 at the frequencies f_1 and f_2, the instantaneous frequency can be calculated to [45], [46]

$$f^{\text{inst}} = \frac{1}{2}(f_2 + f_1) + (f_2 - f_1)\,\frac{A_2^2 - A_1^2}{\hat{A}^2(t)}, \tag{3.30}$$

with

$$\hat{A}(t) = \sqrt{A_1^2 + A_2^2 + 2A_1A_2\cos\left(2\pi(f_2 - f_1)t\right)}. \tag{3.31}$$

Thus, for symmetric excitation, the instantaneous frequency equals the center frequency, such that no difference is observed.

Although the usage of f^{inst} offers some advantages in certain cases, it needs to be handled with care. For instance, when the instantaneous phase has discontinuities or rapid variations, the changes in the resulting frequency could possibly be too fast to comply with the fundamental constraint of slow signal variations. Also, for many communication signals, the instantaneous frequency would oscillate closely around or be approximately equal to a dominant frequency component, e.g., the center frequency. Thus, for the following investigations in this thesis, when no further comment is given, the center frequency approach is chosen as the reference for the model, as it is easy to define for all scenarios and it matches the dominant frequency for equal-amplitude two-tone characteristics.

Interpolation Schemes

The reference data, no matter whether extracted from measurements or simulation, is necessarily limited to a number of sample points. A single-frequency SGC model refers to sampled data in ΔP_{in} steps. In multi-frequency models, as the instantaneous frequency SGC model, additional sampling dimensions must be taken into account. For the model usage an interpolation scheme is required. Different interpolation schemes have been tested. Linear interpolation is the most simple choice, but results in non-differentiable curves and introduces numerical noise into the calculation [47]. In fact, the error is proportional to the squared distance between sample points [48], so it imposes more demanding conditions on the reference data. While this is manageable for the single-frequency SGC model, already for the instantaneous frequency SGC model much more data is required. This issue further increases with model complexity. Thus, other, higher order interpolation schemes are beneficial. For oscillating curves cubic spline interpolation [47] typically offers good results. Here, polynomials are fitted in-between the sampling points, such that not only the resulting interpolated curve, but also the first and second derivative are continuous. In contrast, using a monotone piecewise cubic interpolation [49] does not result in a continuous second derivative. While overshoots and unwanted oscillation might occur in the spline scheme, this is not the case for the piecewise cubic interpolation, at the cost of possible second derivative discontinuities and curvature steps. For the models in this thesis, in most cases piecewise cubic interpolation is preferred and used for both drive- and frequency-based interpolation, as the behavior of the interpolated curves can be considered non-oscillating, and overshoots, e.g., between frequency points, hurt far more than any discontinuity in the second derivative.

Simulation of Effective Efficiencies and Collector Currents

In principle, various output quantities of the TWT can be estimated using Envelope methods. Apart from the output waveform and related characteristics on the signal side, the efficiency itself is of high interest, as its optimization is a key factor to improve TWTs in back-off. As noted before, the envelope of a modulated signal changes over time. Thus, the effective short-time operating point moves across the transfer curves in time. It spans across a range of input power levels, for which an effective[3] efficiency is to be determined.

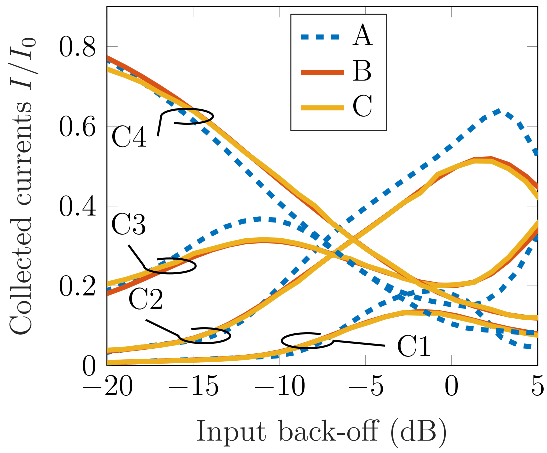
According to Section 2.4, the total efficiency of a TWT can be calculated from the DC power inserted into the system, the beam efficiency, and the regained electric power in the collector. While in the CW case, these quantities can be described by static values over time, the effective values for modulated operation are weighted time-averages. The DC input power as well as the collector stage potentials are fixed known TWT parameters, and the effective beam efficiency can be calculated from the output signal in an Envelope approach as previously described.

In contrast to what can sometimes be read in literature, neither effective collector efficiency nor the effective total efficiency can be directly calculated by averaging the respective steady-state efficiencies with the probability of occurrence of the respective input power level over time. This is, because the collector efficiency relates the regained electric power to the total power entering the collector. Both quantities in this ratio change over time according to the input signal envelope distribution, which forbids a direct envelope averaging scheme, e.g., using the SGC model, on $(P_{\mathrm{in}}, \eta_{\mathrm{Coll}})$ as the reference data. Similarly, this holds for the total efficiency. Instead, a detour over other quantities, for instance the currents in the individual collector stages, has to be taken.

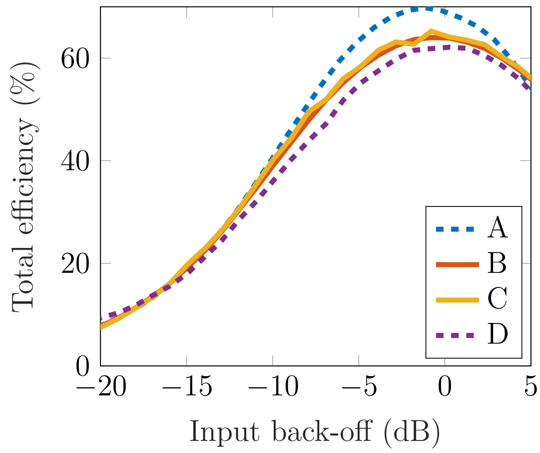
In Figure 3.9, the estimation of the effective total efficiency is compared to measurements[4]. In Figure 3.9a, the effective currents of a four-stage collector are calculated by SGC models of the individual collector stages. The SGC models relate the effective

[3] In the following, an effective metric describes the expected time-averaged metric at the output of the TWT from a modulated input signal.

[4] The definition of the term input back-off is not consistent in literature. In this work, the input back-off levels are defined by $P_{\mathrm{in}}/P_{\mathrm{in,sat}}$. This may feel wrong from a linguistic point of view, but enables presenting the results in the most intuitive way, with low powers to the left and high powers to the right. The same holds for the output back-off.



(a) Collector currents at the stages C1 to C4.



(b) Total efficiency.

Figure 3.9: SGC simulation of collected currents and the total efficiency for 16QAM signals based on measured characteristics. (A): CW measurements. (B): 16QAM measurements. (C): 16QAM simulation with SGC models of the collector stages. (D): 16QAM simulation with an SGC model of the total efficiency as the output characteristic.

currents to reference CW measurements of the collected currents per stage over a range of input back-off levels, from small-signal to overdrive operation. For a random 16QAM signal, the resulting effective currents per stage are calculated and compared to measured results for a 16QAM signal excitation. As can be seen, the agreement is excellent over most of the considered power level range, with some deviation visible on the lower end. On the right side, in Figure 3.9b, the total efficiency as calculated from the currents in Figure 3.9a is compared to 16QAM measurements of the total efficiency. It can be seen, that in contrast to a direct estimation based on the total efficiency as a reference, the detour allows accurate estimation of the effective total efficiency in modulated operation of the TWT.

Effective Electron Velocity Spectrum Simulation

In Section 3.3.2, the effective efficiency is estimated based on measured collector currents. In simulation, the currents must be calculated first from the spent-beam EVS. As can be seen from Figure 2.3, the spectral properties vary strongly with drive, such that the collector stage potential choice significantly influences the collector efficiency, and thus the total efficiency. Therefore, for fully optimizing a TWT in modulated back-off, it is worthwhile knowing the effective EVS. In Figure 3.10, the estimation approach for this effective spectrum is shown for an exemplary two-tone signal. For each drive level, the spent electron beam from steady-state simulation in MVTRAD can be evaluated regarding its velocity spectrum, resulting in an array of curves as shown in Figure 3.10c. Taking the full envelope spread into account, the array can be used as a reference for a two-dimensional SGC model. With this, the weighted time-averaged macro-particle energy distribution can be calculated, which can be seen in Figure 3.10d. For comparison, the spent beam from a direct physics-based two-tone solution in MVTRAD is added. It can be seen, that the SGC simulation based on MVTRAD CW results enables accurate estimation of the modulated EVS.

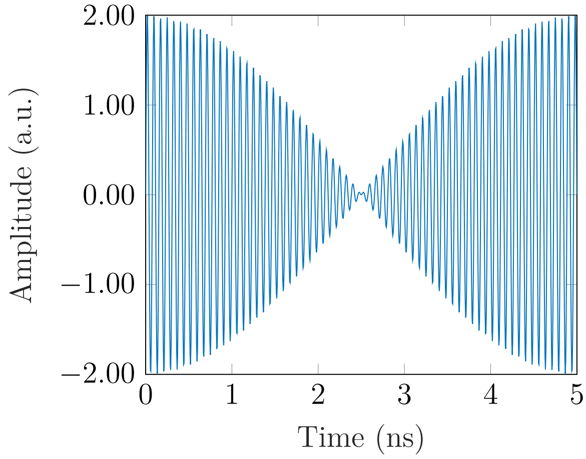
Model Schematic and Signal Flow

The resulting hybrid SGC model is summarized in Figure 3.11. The input signal $x(t)$ is evaluated regarding its frequency components, which is used for the frequency reference choice. The TWT's steady-state information is generated in ΔP_{in} steps for one or multiple frequencies, e.g., in MVTRAD. After interpolation, the respective SGC models for the output signal, EVS or other desired properties, for instance the i-th collector stage current I_{c}^{i}, are generated. From the Hilbert transform of the signal $x(t)$, the preprocessed analytic signal $x^{\text{inst}}(t)$ is calculated, which is inserted into the SGC models to calculate the output quantities.

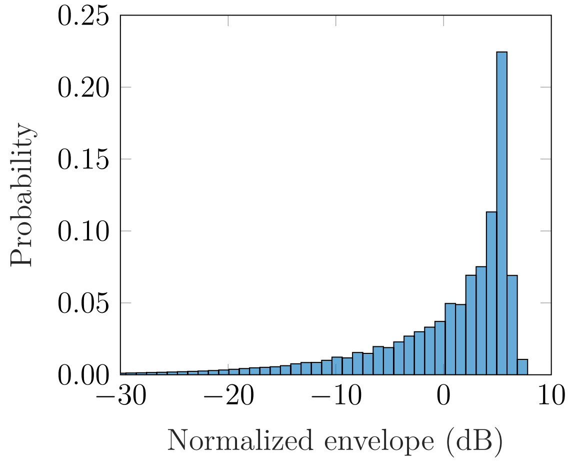
3.3.3 Frequency-Dependent Model Topologies

As previously established, the dispersive behavior of the TWT stems from that of the delay line. By means of tapering techniques, geometric adaptions of the line, or even special topologies, the "dispersiveness" can be mitigated, such that for classical, simple modulations on small bandwidths (compared to the frequency-domain variations of the amplifier), frequency-independent Envelope methods are expected to suffice. Still, the limits of such methods are not inherently clear. Especially future topologies for higher frequencies, e.g., FW lines, possibly impose higher requirements to tools for estimating the broadband behavior of the TWT.

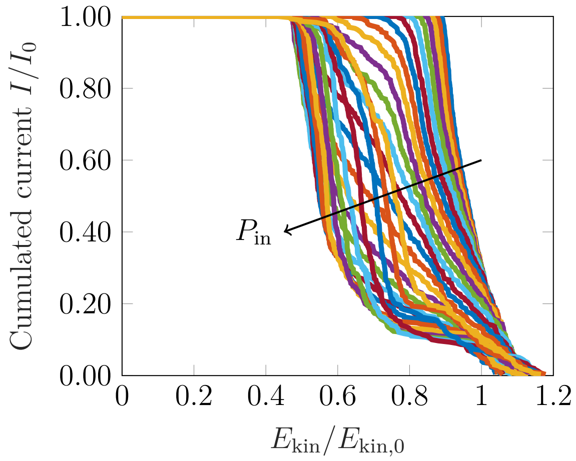
But already today, in the case of more broadband excitation, the fundamental Envelope approaches have inherent limitations in terms of the usable bandwidth to avoid dispersion-based distortions of the resulting signal. Thus, in the past, various more involved approaches have been formulated [50]–[52]. They offer the inclusion of the frequency-dependence, which is inherent to the beam-wave interaction, at the cost of strongly differing model structures and more complex implementation and usage. As the general idea behind this work is to generate models that can be used,



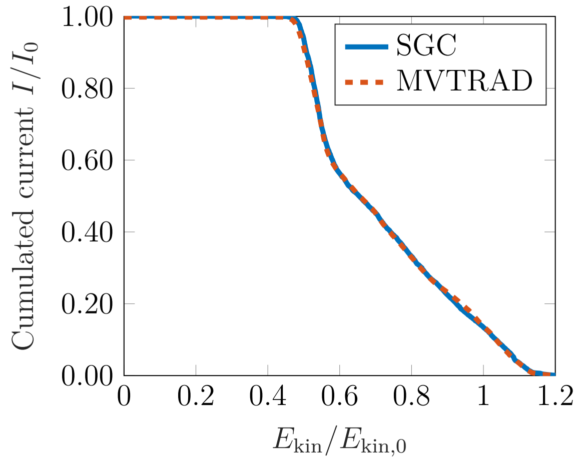
(a) Modulated two-tone signal.



(b) Amplitude distribution occurrence probability.



(c) CW EVS for various power levels.



(d) EVS from the two-tone signal.

Figure 3.10: Estimation of the velocity distribution of the spent-beam current.

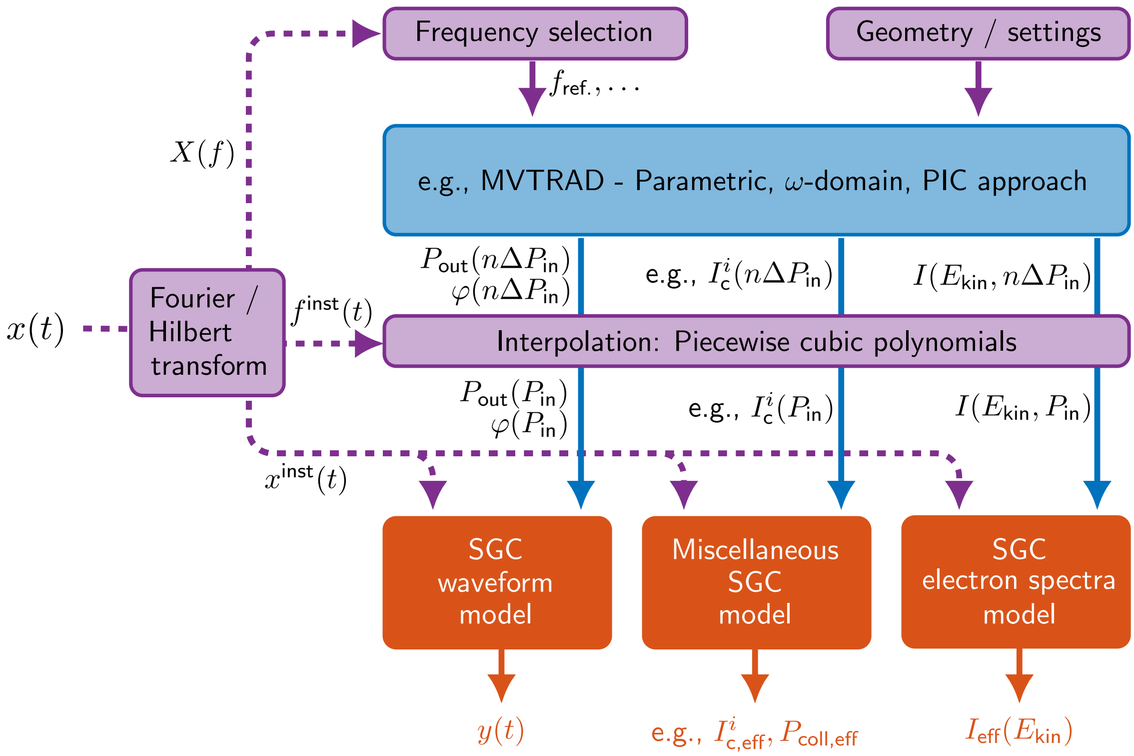


Figure 3.11: Schematic of the hybrid SGC model.

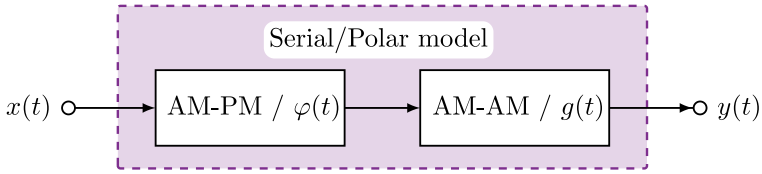


Figure 3.12: Serial model topology.

e.g., already at the design stage, possibly for optimization of a TWT, not all models found in literature are suitable. Most models are developed with the aim to accurately depict the behavior of a known amplifier, for instance for a system analysis containing the fixed amplifier, or for the design of peripherical components, e.g., linearizers or filters. For these applications, today, there basically are no significant limits to the complexity of the fitting process or the reference data to generate the model. One could for instance use measured, complexly modulated signals to ensure that the model will precisely predict this reference data.

A selection of models from literature is described in [53]. The models can roughly be sorted into three categories: models in serial structure (as shown in Section 3.3.2 and Section 3.3.4), models in quadrature structure (as will be described in Section 3.3.5) and models which do not fit into the two categories. The model type selection and terminology is adapted from and inspired by [53], where approaches for the modeling and simulation of nonlinear systems are outlined.

Serial or polar models are following the most intuitive order to describe the amplifier, as shown in Figure 3.12. In principle, the nonlinearities of the transfer curve can be considered as two seperate phenomena. The first one represents the phase distortion, i.e., the AM-PM effect, characterizing the phase shift observed when driving the amplifier to a certain operating point. The second one describes the amplitude compression, i.e., the AM-AM effect. In a serial model, those two mechanisms are directly applied in one way or another. One example is the frequency-independent model as given in Section 3.3.2. There are certainly more complexly structured models available in literature, following the same serial topology.

A well known but sometimes fairly involved group including frequency-dependent components is the Multi-Box serial model family. In these models, individual effects are summarized by separate functional blocks, which are connected to form the full model. As most polar models, they inherently assume a shape- and therefore functional similarity of the respective AM-AM and AM-PM transfer curves over frequency. In many models, when merging the dispersive behavior of the TWT with the drive-dependent characteristics into one representation, the two effects are still calculated separately and concatenated afterwards. As a result, with the analytic signal representation $x_0(t)$ and the corresponding instantaneous phase $\xi_0(t)$ of the input signal $x(t)$ (as defined by Equations (3.24) to (3.26)), the output signal of the TWT can typically be described as

$$y(t) = g_{\mathrm{AM}}\left(|x_0(t)|\right) \cdot \exp\left(j\xi_0(t) - \Theta_{\mathrm{PM}}\left(|x_0(t)|\right)\right). \tag{3.32}$$

with

$$g_{\mathrm{AM}} := g_{\mathrm{AM}}^{(\omega)}(\omega, A) \tag{3.33}$$

and

$$\Theta_{\mathrm{PM}} := \Theta_{\mathrm{PM}}^{(\omega)}(\omega, A). \tag{3.34}$$

In contrast to Equation (3.27), here the amplitude and phase modifiers $g_{\mathrm{AM}}^{(\omega)}$ and $\Theta_{\mathrm{PM}}^{(\omega)}$ are functions of amplitude and frequency. As frequency itself is defined as the number of cycles over time, we cannot determine a literal instantaneous frequency at any instance. Any frequency-dependence therefore requires some sort of memory in the model. This is why a model as described in Section 3.3.2 is sometimes referred to as memoryless, even though this is not entirely correct in the presented implementation, as any envelope-based approach inevitably is a memory-based approach in itself. In frequency-dependent models, the memory serves to detect the spectral information of the signal in addition to the envelope and phase as required by the SGC model. This enables filtering of the signal according to the model properties and the short-time frequency content of the signal. In general, filtering is conceivable both before and after the nonlinearity, resulting in a now altered equation for the output signal

$$y(t) = h_{\mathrm{out}}(t) * \left[g_{\mathrm{AM}}\left(|x_1(t)|\right) \cdot \exp\left(j\xi_0(t) - \Theta_{\mathrm{PM}}\left(|x_1(t)|\right)\right)\right], \tag{3.35}$$

with

$$x_1(t) = \left(h_{\mathrm{in}} * x_0\right)(t). \tag{3.36}$$

The two filters h_{in} and h_{out} represent different aspects of the TWT's behavior and cannot be merged or interchanged. Simplified, the first one predominantly incorporates drive variations over frequency by considering the flatness of the input back-off, while the second one in addition predominantly readjusts the spectral response in the nonlinear regime. The actual filter properties need to be obtained in a possibly complicated fitting process. This can be done either in frequency- or in time-domain, leading to a natural description of the resulting filters in the respective domain. Oftentimes, as arbitrarily modulated transient signals are used for fitting, a time-domain approach is more intuitive and more typically applied. In our case, as the idea is to connect the Envelope approach as a module to a hybrid time- and frequency-domain simulation routine, a frequency-domain approach is more natural.

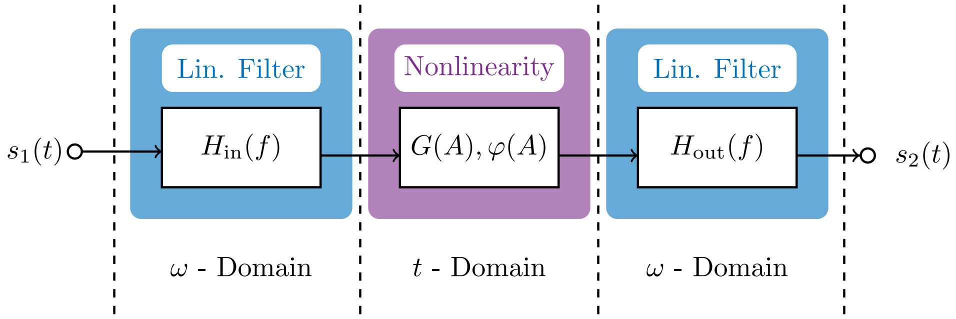


Figure 3.13: Typical chain of nonlinear and frequency-dependent models in a Three-Box model.

3.3.4 Three-Box Model

A natural way of separating and concatenating amplitude- and frequency-dependent parts of the model is the so-called Three-Box model (TBM). Its basic idea is to use the similarity of the AM-AM and AM-PM curves over frequency. The TBM is schematically shown in Figure 3.13. As the name suggests, the functionality is summarized in three functional blocks. The nonlinearity box in the middle contains the nonlinear phase and amplitude characteristics and is naturally applied in time-domain. This is the actual Envelope simulation here. The filters before and after this part representing the nonlinearity are purely linear and therefore do not need to live in time-domain only. They are described and applied in frequency-domain. This requires multiple discrete Fourier transforms to switch between time- and frequency domain, which at carrier level might be rather time consuming.

The operation principle of this model assumes that a change in frequency results in a simple shift of the nonlinear transfer curves both in domain and co-domain of the nonlinear characteristic. Thus, it is assumed that the shape itself does not change significantly over frequency.

This leads to a fairly simple and straightforward filter generation scheme. For an AM-AM shape function $g_{\text{AM},\omega_0}(A)$ at a reference frequency ω_0, the filter $H_{\text{A,I}}$ at the model input can be generated as

$$H_{\text{A,I}}(\omega) := \sqrt{\frac{P_{\text{in, sat}}(\omega)}{P_{\text{in, sat}}(\omega_0)}}, \tag{3.37}$$

with the saturated input power over frequency $P_{\text{in, sat}}(\omega)$, while the respective filter $H_{\text{A,O}}$ at the model output is

$$H_{\text{A,O}}(\omega) := \sqrt{\frac{P_{\text{out, sat}}(\omega)}{P_{\text{out, sat}}(\omega_0)}}, \tag{3.38}$$

with the saturated output power over frequency $P_{\text{out, sat}}(\omega)$. Similarly, for a AM-PM shape function $\Theta_{\text{PM},\omega_0}(A)$, the filters $H_{\text{P,I}}$ and $H_{\text{P,O}}$ at input and output can be built as

$$H_{\text{P,I}}(\omega) := \sqrt{\frac{P_{\text{in, ref}}(\omega)}{P_{\text{in, ref}}(\omega_0)}}, \tag{3.39}$$

and

$$\begin{aligned} H_{\text{P,O}}(\omega) &:= \exp\left[j\,\Theta_{\text{PM},\omega}\left(P_{\text{in, ref}}(\omega)\right) - j\,\Theta_{\text{PM},\omega_0}\left(P_{\text{in, ref}}(\omega_0)\right)\right] &(3.40)\\ &= \exp\left[j\,\varphi_{\text{out, ref}}(\omega) - j\,\varphi_{\text{out, ref}}(\omega_0)\right]. &(3.41)\end{aligned}$$

The tupel $(P_{\text{in, ref}}(\omega), \varphi_{\text{out, ref}}(\omega))$ denotes a reference point on the respective AM-PM transfer curve. In principle, the chosen reference point only has the constraint that the AM-PM transfer curve shapes should be as similar as possible around this point for the desired frequency range. For

$$P_{\text{in, ref}}(\omega) = P_{\text{in, sat}}(\omega) \tag{3.42}$$

or at least

$$\frac{\delta P_{\text{in, sat}}}{\delta P_{\text{in, ref}}} = 1, \tag{3.43}$$

the filters as described in Equation (3.37) to Equation (3.41) can directly be used in a Three-Box scheme. Both the order in which the filters and nonlinearities are aligned and the filter synthesis are schematically shown in Figure 3.14. The nonlinearity-block containing the AM-AM and AM-PM transfer function can be implemented as shown in Section 3.3.2. If Equation (3.43) does not hold, the filters need to be reassorted. By this, the resulting scheme is identical to a more general Multi-Box model, which is sometimes referred to as Poza-Sarkozy-Berger (PSB) model [52]. For this, the nonlinearity block itself consists of an AM-PM nonlinearity, a following filter

$$H_{\text{PSB}}(\omega) := \frac{H_{\text{A,I}}(\omega)}{H_{\text{P,I}}(\omega)}, \tag{3.44}$$

and afterwards the AM-AM nonlinearity. This block is then framed by two linear filter blocks. The preceding filter is

$$H_{\text{PSB,in}}(\omega) := H_{\text{P,I}}(\omega) \cdot H_{\text{P,O}}(\omega), \tag{3.45}$$

and the subsequent one is

$$H_{\text{PSB,out}}(\omega) := H_{\text{A,O}}(\omega). \tag{3.46}$$

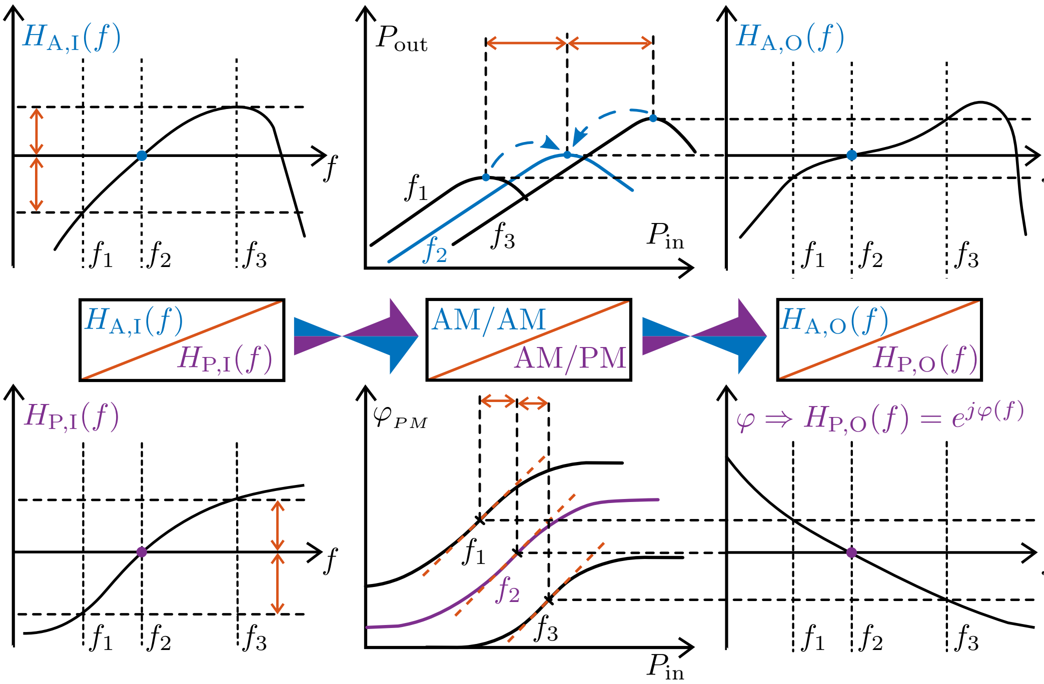


Figure 3.14: Schematic view of the TBM, adapted from [53].

This representation offers more flexibility regarding the reference point than the TBM, but is more complex and not modular with respect to the SGC model.

As noted before, the TBM assumes a static transfer-curve shape over frequency. This is of course a rather limiting simplification, as only curves in a certain limited back-off range can be seen as similar. The range shrinks even further, the larger the frequency shift is. Therefore, only limited bandwidths can be represented accurately in the model.

Still, in some cases the model improves the estimation compared to a simple frequency-independent model, as is shown in [54]. Here, the average D_3 for an equal power two-tone excitation on a Ku-Band TWT is evaluated. The reference TWT model used in [54] is more dispersive than typical commercial Ku-Band TWTs. Thus, all frequency-dependent effects are emphasized. Results from a TBM according to Equation (3.42) are compared to reference MVTRAD simulation. The findings show that while for some narrow frequency spacings, the TBM improves the estimate compared to the SGC models, it is still limited in bandwidth and in many cases does not offer much improvement compared to the frequency-dynamic SGC model representation.

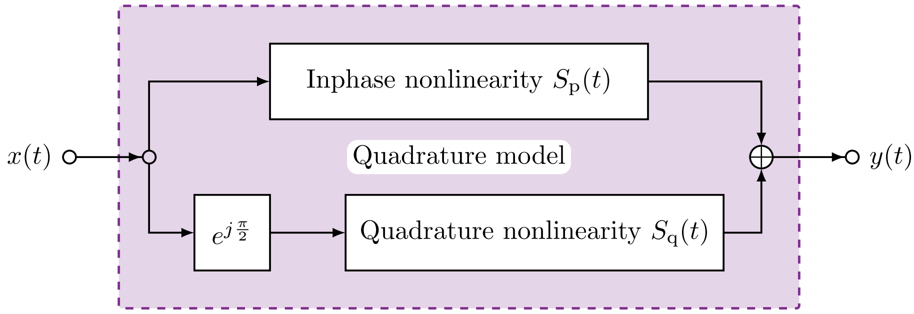


Figure 3.15: Typical structure of models in quadrature topology according to [53].

Thus, the TBM has both advantages and disadvantages compared to the SGC approaches. To some extent, it mitigates the issue of the ambiguous frequency reference choice that comes with the SGC models, adding some frequency information to the estimate. Unfortunately, this comes at the cost of a limited back-off range that can be considered. Also, the similarity of the AM-AM and AM-PM curves over frequency might vary strongly depending on the TWT, such that the scope of usability of such a model is strongly reduced. Especially when the TWT is strongly dispersive, it is more important to use a model which is more broadband than the SGC model, but typically, this also entails a more distorted curve shape. Thus, one needs to be very careful when using either the TBM or also the PSB model, which has similar constraints.

3.3.5 Quadrature Polynomial Model

General Idea and Inspiration

While the TBM and similar serial models improve the estimation for broadband signals, they have unfavorable properties, as discussed in the previous section. Especially the shape-similarity constraint strongly limits the usability.

Thus, more involved models are considered. From the variety of models, there are several ones with a high degree of flexibility. In principle, one could arbitrarily concatenate blocks of nonlinearities and linear filters to render the TWT's properties. Then, a large variety of individual functional blocks leads to highly adaptable structures with many degrees of freedom. The parameters corresponding to the considered device then need to be fitted, if possible, to as many input signal constellations as possible, as for instance required in a general Volterra [53] approach. Building such a model based on an existing device, which can be characterized based on arbitrary signal waveforms, one can expect the outlined approach to yield correct results at the cost of being tedious and time-consuming.

Our problem is different. Instead of being able to reproduce the behavior of the amplifier with some accuracy, the idea here is to predict it for multi-frequency excitations, solely based on information effortlessly obtained by, e.g., frequency-domain simulation results or classical measurements. This leads to a confinement of choices from the pool of models. A subgroup of models can be based on sets of single-tone characteristics. The frequency-dependent nonlinear amplifier model can be constructed by splitting it into a number of concatenated blocks, following the description in Section 3.3.3. The evaluation of the non-dispersive nonlinearities is carried out in time-domain and the corresponding linear parts are represented by bandpass filters which are evaluated in frequency-domain. A well known and rather simple type of model from the usable subgroup is the so-called Saleh model [50]. It is best known as a single-frequency approach, where the transfer curve is fitted by an expression with two parameters for the amplitude and and two more for the phase information. Additionally, in [50], a frequency-dependent model is proposed, where the characteristics are considered in two distinct parts, an inphase branch and a quadrature branch with a 90° phase shift. The idea of this separation stems from resorting Equation (3.32) by Euler's formula, resulting in the quadrature description [55], as can be seen in Figure 3.15. Both the inphase and quadrature part are then fitted to two analytic expressions, each with two parameters per considered frequency point. Thus, the model requires only four degrees of freedom to describe the characteristics of the amplifier at a single frequency. Therefore, the model identification is as simple as it gets. Unfortunately, for the purpose defined in this work it has shown to be insufficient to predict the nonlinearities and the dispersiveness of the TWT with the required accuracy.

A more versatile model, enabling a closer fit, is the so-called Bessel function model (BFM) [51]. For this, the expressions for the inphase and the quadrature components are generalized to a sum of Bessel functions. Instead of being limited to four, the number of parameters is now a parameter itself. The adaptability comes at the cost of a complex topology and fitting mechanism. In total, K selected gain and phase curves over frequency are translated to the respective inphase and quadrature components S_{I} and S_{Q} of the transfer characteristics, given by

$$S_x(A, f) = g_{\mathrm{AM}}(A, f) \cdot \begin{cases} \cos(\Theta_{\mathrm{PM}}(A, f)), & x = I, \\ \sin(\Theta_{\mathrm{PM}}(A, f)), & x = Q. \end{cases} \tag{3.47}$$

These components are then each fitted by N Bessel functions of first order and kind

$$S_{\mathrm{I,Q}}(A, f) \overset{!}{=} \sum_{n=1}^{N} G_{\mathrm{I,Q}}(n, f) \cdot J_1\left(\frac{n\pi A}{D}\right), \tag{3.48}$$

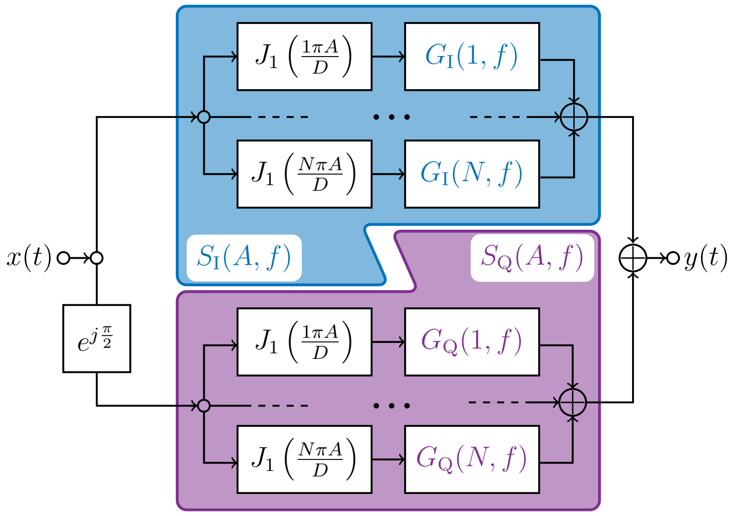


Figure 3.16: Simplified block schematic of the BFM.

with a shape-controlling parameter D of the individual Bessel functions and $2 \cdot N \cdot K + 1$ degrees of freedom. The quadrature and the inphase components contain N branches each. These branches, in turn, are each composed of a Bessel function and a linear filter according to Equation (3.47). The parameters $G_{\mathrm{I,Q}}(n, f)$ are sorted and reshaped into $2N$ linear filters, which are applied subsequently. A simplified schematic of the BFM is shown in Figure 3.16. Simulation has shown that the number of Bessel branches is required to be above 20 for reasonable results.

As has been noted, one drawback, amongst others, is the model complexity. Thus, a simpler model inspired by the BFM and the other principles given in the previous sections is derived. Again, the TWT is split into a quadrature and an inphase component, as this has shown to be useful for mitigating the shape-similarity constraint of polar models. Still, one could imagine a similar approach in a serial way. Thus, the composition of the quadrature and the inphase components is the more crucial part. From other models, e.g., general Volterra or Polyspectral approaches, the separation of the transfer function into polynomial partial functions is well-known. Therefore, inspired by these and in combination with the quadrature approach, each quadrature branch is split into partial functions containing a part of the total nonlinear dispersive behavior. A signal going through, for instance, the inphase branch then sees a different part of the total nonlinearity in each subbranch. These partial nonlinearities are constituted by partial polynomial terms. The nonlinear blocks are then evaluated in time-domain with the envelope extracted using the Hilbert transform.

Structure and Workflow of the Quadrature Polynomial Model

The structure of this Quadrature Polynomial Model (QPM) is presented in Figure 3.17. First, the reference data is generated in the MVTRAD part, based on a 2D grid of input powers $[P_{\text{in,min}}, (P_{\text{in,min}} + \Delta P_{\text{in}}), \ldots, P_{\text{in,max}}]$ and frequency points $[f_{\min}, (f_{\min} + \Delta f), \ldots, f_{\max}]$. The values $P_{\text{in,min}}$, $P_{\text{in,max}}$, $f_{\min}$, and $f_{\max}$ are the lowest and highest considered input powers and frequencies, respectively. The grid points are equally spaced with ΔP_{in} and Δf. For each combination, an MVTRAD run is evaluated, such that the output power and phase are obtained and stored. This 2D data is translated to a two-dimensional piecewise cubic polynomial (see Section 3.3.2), which is then considered in the fitting routine used for the model parameter identification. A further explanation follows in the following section.

The Envelope model itself is split into two parts which are in quadrature. Each block consists of N branches according to the chosen number of degrees of freedom. Each branch contains a TBM according to Section 3.3.4, with Equation (3.43), where the model parameters from the fitting routine describe the TWT behavior. The individual TBM-blocks represent different parts of the transfer characteristics. Thus, the TBM filters describe the impact of the partial nonlinearity considered in the block. In principle, the nonlinear base functions used for the expansion can be chosen in many ways. Here, monomial terms of varying order are used, as they have shown to be well manageable. In principle, one can imagine several satisfactory solutions based on the degrees of freedom and base function description. Similarly, different but nevertheless still well-fitting parameters can be found without a frequency-dependence of the output filter in each branch. This does not directly comply with the intuitive description in Section 3.3.4, as the model is not based on a direct intuitive relationship between physical properties and the model parameters. Nevertheless, an attempt to find such a relationship in the QPM will fail anyway. Thus, for the sake of a reduced model complexity, for the further study and simulation results in this work $G^2_{\text{I,Q}}(n, f)$ is set to one.

According to Figure 3.13, the filters are applied in frequency-domain, followed by a time-domain evaluation of the partial nonlinearities. Therefor, the nonlinearities are each considered in an individual SGC model with a Hilbert transform. To save computation time, the Hilbert transform is merged with the inverse-FFT, as it is required anyway. Also, as the Fourier transform is linear, an FFT is performed before the signal enters the branch-points in front and behind the inphase and quadrature components of the model, respectively. For the output signal $y(t)$, the summed resulting signals from the branches from both inphase and quadrature part are superimposed.

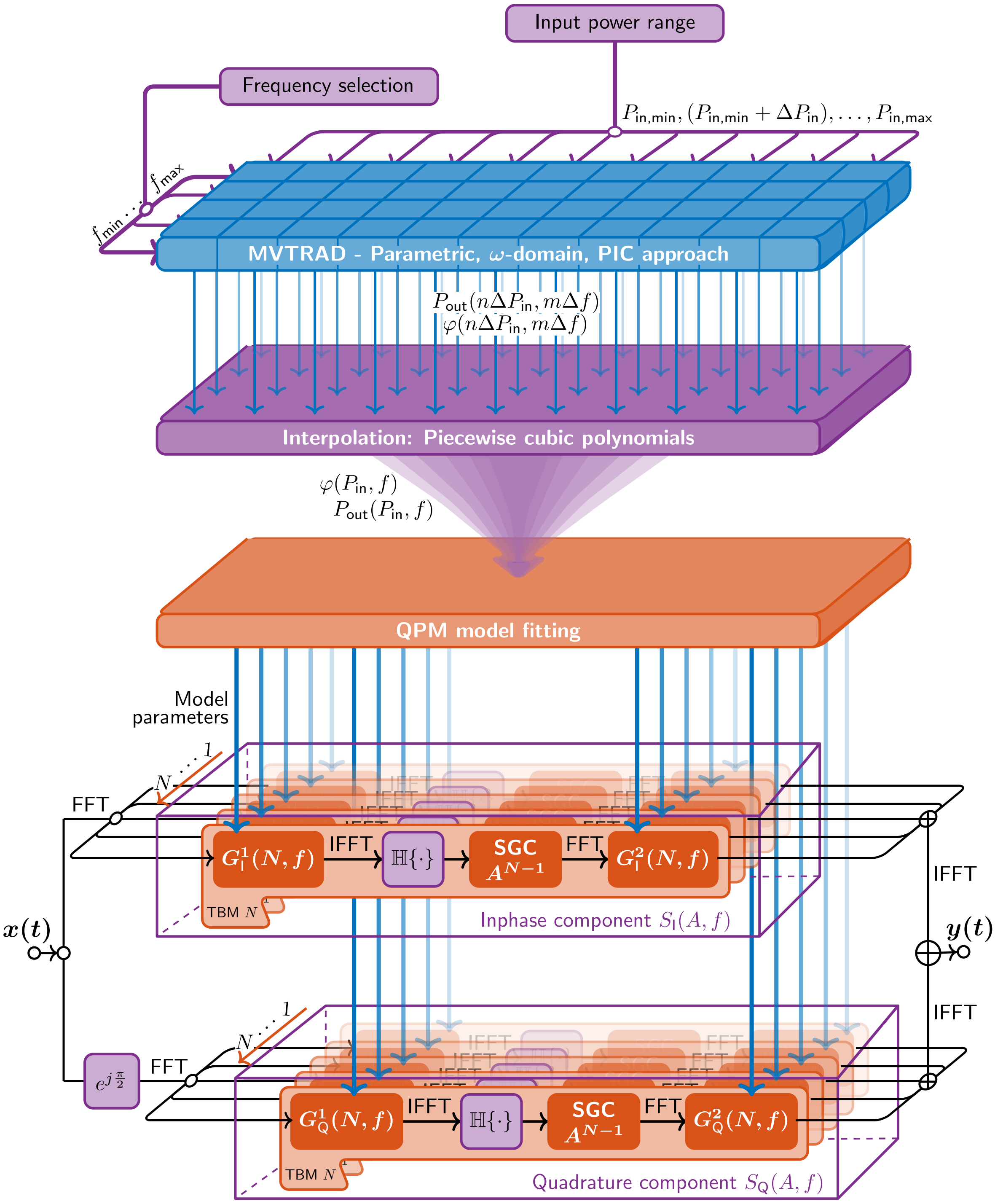


Figure 3.17: Full topology of the QPM and the model identification scheme.

Model Identification and Fitting Routines

For accurate simulation with the QPM, the model parameters need to be well-fitted to the reference data as generated in MVTRAD. In the following, suitable fitting routines are established. The general idea is, that

$$\min_{G_{\mathrm{I,Q}}} \left[\sum_{f,A} \left| S_{\mathrm{I,Q}}(A,f) - \sum_{n=1}^{N} G_{\mathrm{I,Q}}(n,f) \cdot q_{\mathrm{I,Q}}\left(n,A\right) \right|^2 \right] \tag{3.49}$$

holds, where $q_{\mathrm{I,Q}}$ describes the base functions. In the special case described in this section and Figure 3.17, these base functions relate to the base monomials A^n.

With a vector of input amplitudes $\mathbf{A}(f_i) = [A_1, A_2, ..., A_M]$, possibly interpolated from the grid points, and the respective output inphase and quadrature components $\mathbf{S}_{\mathrm{I,Q}}(f_i)$, the required parameters can be found as

$$\mathbf{G}_{\mathrm{I,Q}}(f_i) = [\mathbf{M}]^{+}(f_i)\,\mathbf{S}_{\mathrm{I,Q}}(f_i), \tag{3.50}$$

where $[\mathbf{M}]^{+}(f_i)$ is the Moore-Penrose pseudoinverse [56], [57] of the transfer matrix

$$[\mathbf{M}](f_i) = \left[\mathbf{A}^0(f_i), \mathbf{A}^1(f_i), \cdots \mathbf{A}^{N-1}(f_i)\right]. \tag{3.51}$$

Here, $\mathbf{A}^n$ denotes the elementwise Hadamard product [58]

$$\mathbf{A}^n = \mathbf{A} \circ \mathbf{A}^{n-1}. \tag{3.52}$$

The Moore-Penrose pseudoinverse is calculated by estimating the singular value decomposition of $[\mathbf{M}]$

$$[\mathbf{M}] = U\,[\mathbf{\Sigma}]\,V^*, \tag{3.53}$$

with the diagonal matrix

$$[\mathbf{\Sigma}] = \operatorname*{diag}_{i}\{\sigma_i\} \tag{3.54}$$

and σ_i denoting the singular values of $[\mathbf{M}]$. For the numerical evaluation, singular values below a threshold ϵ are not considered, i.e., set to zero. Thus, with the dimension-wise reduced matrix

$$[\mathbf{\Sigma}_\epsilon] = \operatorname*{diag}_{\sigma_i > \epsilon}\{\sigma_i\} \tag{3.55}$$

and the accordingly truncated matrices U_ϵ and V_ϵ, the pseudoinverse is

$$[\mathbf{M}]^{+} \approx V_\epsilon [\mathbf{\Sigma}_\epsilon]^{+} U_\epsilon^*, \tag{3.56}$$

with

$$[\mathbf{\Sigma}_\epsilon]^{+}_{i,i} = \begin{cases} \frac{1}{\sigma_i}, & \text{for } \sigma_i \neq 0, \\ 0, & \text{for } \sigma_i = 0. \end{cases} \tag{3.57}$$

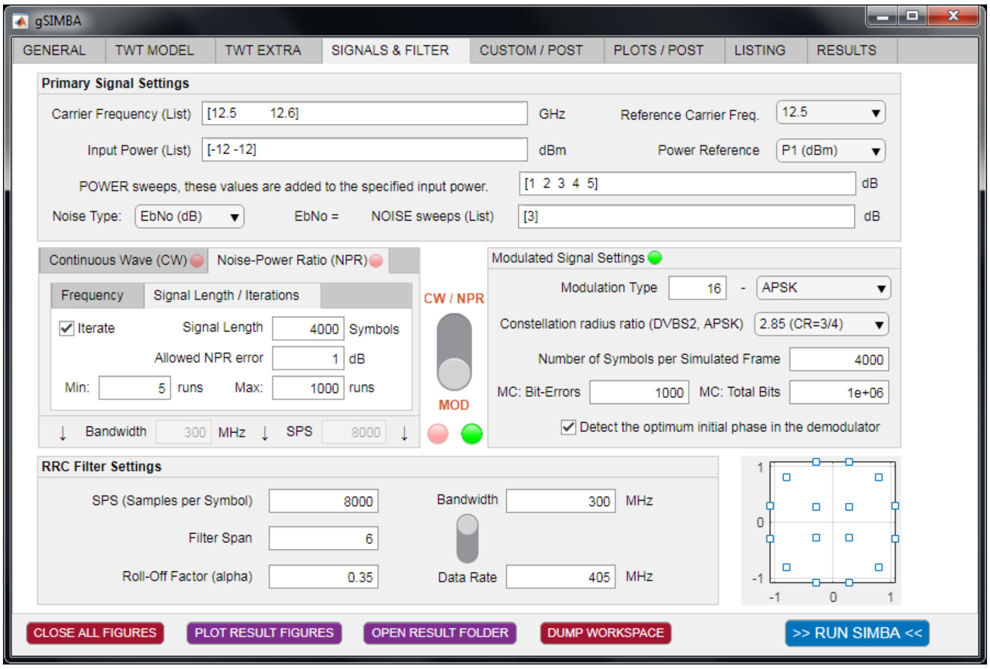


Figure 3.18: SIMBA program written in MATLAB.

With $M = N$, the pseudoinverse equals the inverse and an explicit formulation is available [59]. For $M > N$, there is a unique pseudoinverse, which returns the best solution to the least-squares problem. Thus, the QPM inherently features an explicit best fit for a specified number of degrees of freedom. Also, the fitted parameters have shown to be less dynamic compared to the BFM, such that interpolation, e.g., between frequency points is less critical.

3.3.6 SIMBA and the Satellite Downlink Chain

In order to use the developed models reproducibly, the so-called SIMBA (***S**imulation of **I**nter-**M**odulations in **B**ack-Off and **A**pplications*) program viewed in Figure 3.18 has been developed. Just as the Envelope models themselves, the interface is written in an object-oriented manner in the MATLAB programming language and can be extended modularly.

The amplifier models are integrated in a satellite downlink chain model including modulation and demodulation of communication signals. The block schematic of the signal chain as implemented in SIMBA is shown in Figure 3.19. For the TWT, both the SGC model and the QPM are included. In addition, as it is typically required in such a downlink chain, a linearizer can be inserted ahead of the TWT model. Here, for the sake of simplicity, it is described in an SGC manner. In principle, a combined linearizer and TWT model could be generated, e.g., in a QPM manner, but for some applications it is worth considering them as two separate subsystems. This enables visualizing the predistortion behind the linearizer. Also, when considering the total efficiency, and especially the collector efficiency, the collected currents only depend on the signal distribution at the input of the TWT, which obviously differs from the one at the linearizer input. The collected currents, efficiencies, and electron spectrum estimations are included as a part of the amplifier model blocks. Also, both in front of and behind the TWT, white Gaussian noise can be added to the signal.

Supported signal types are single- and multi-tone signals, typical (single- or multi-carrier) modulation schemes as defined in the DVB-S2 standards, and band-limited, noise-like signals, as are for instance required for noise-power ratio (NPR, [60]) calculations. Communication signals are generated from binary random data, which is plugged into the modulator. Possible modulation schemes are quadrature phase-shift keying (QPSK), M-ary quadrature amplitude modulation (MQAM), M-ary phase-shift keying (MPSK) and M-ary amplitude- and phase-shift keying (MAPSK), with typical constellation radii ratios. The baseband signal generated by the modulator is upsampled and scaled to the desired power level. A Root-Raised Cosine (RRC) filter [61] is implemented for pulse-shaping. At the receiving end, a second Root-Raised Cosine filter is used as a matched filter. Based on the demodulator following the matched filter and the downsampling block, various communication system performance characteristics are calculated, such as bit- and symbol-error rates and the error-vector magnitude.

The models are applied at the carrier frequency, as this is the natural, physical setting. In principle, one could stay at the baseband frequencies, but that would limit the scope of investigations considered in this work.

3.4 Comparison

In the following, the presented QPM and SGC model are evaluated regarding their performance in multi-tone operation. The BFM is also investigated as a benchmark, as it follows similar principles with comparable complexity as the QPM. The material for this section draws in parts on previously published studies [43], [44].

Here, the comparison is carried out in several steps. As the main objective of the considered approaches is to enable the simulation in the design process, the models are not supposed to be based on measured TWT characteristics, which is different to many other approaches. Still, in the first step, measurements are used as a reference,

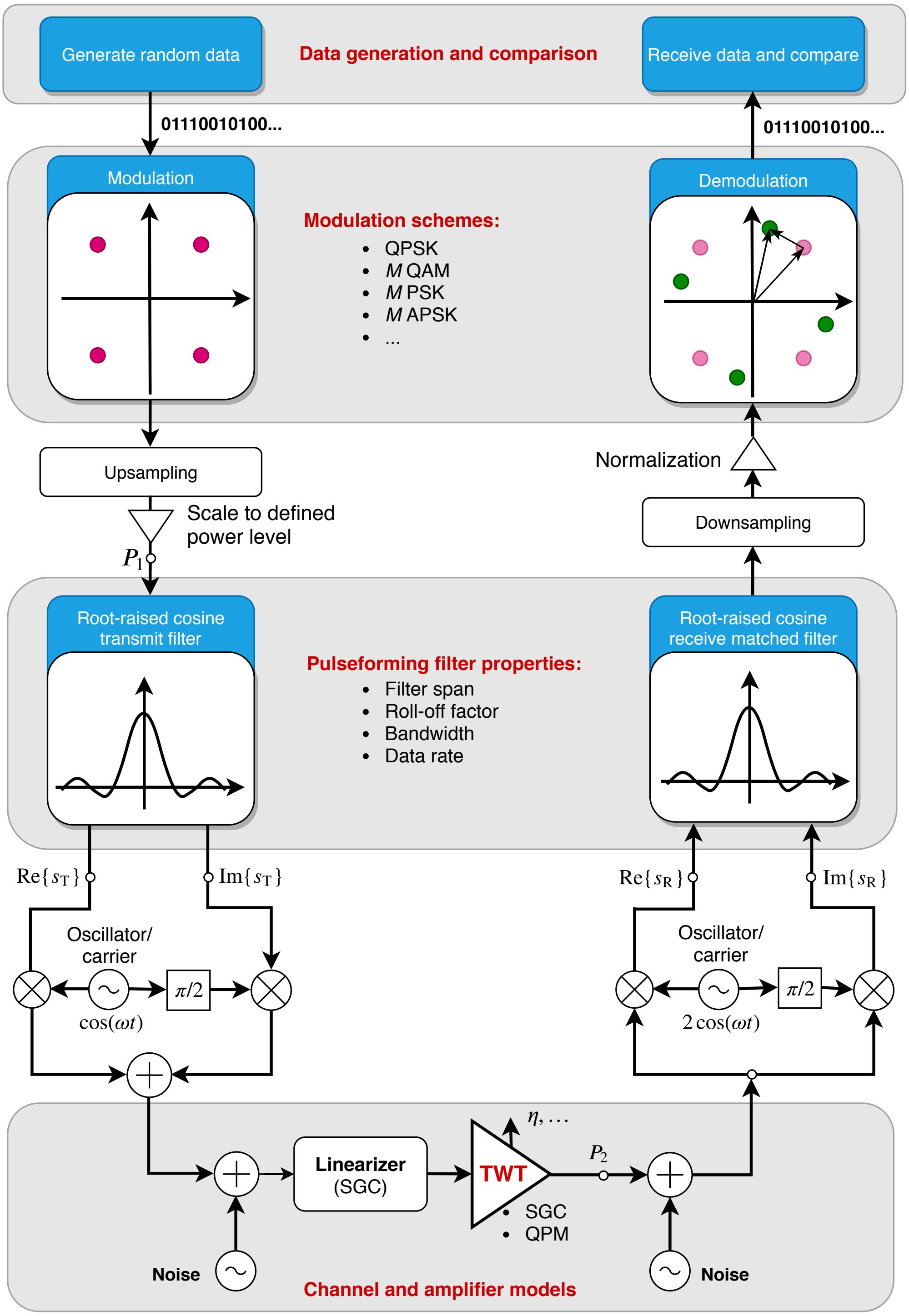


Figure 3.19: Schematic view of the signal chain from a single run with a communication signal in SIMBA.

as this mitigates other error sources, such as the numerical implementation or other inaccuracies in the underlying interaction model.

Then, in a second step, the measured data is replaced by that from the steady-state simulation tool MVTRAD. This defines the combined hybrid frequency- and time-domain approach, which consists of the steady-state simulation of MVTRAD and the presented Envelope codes.

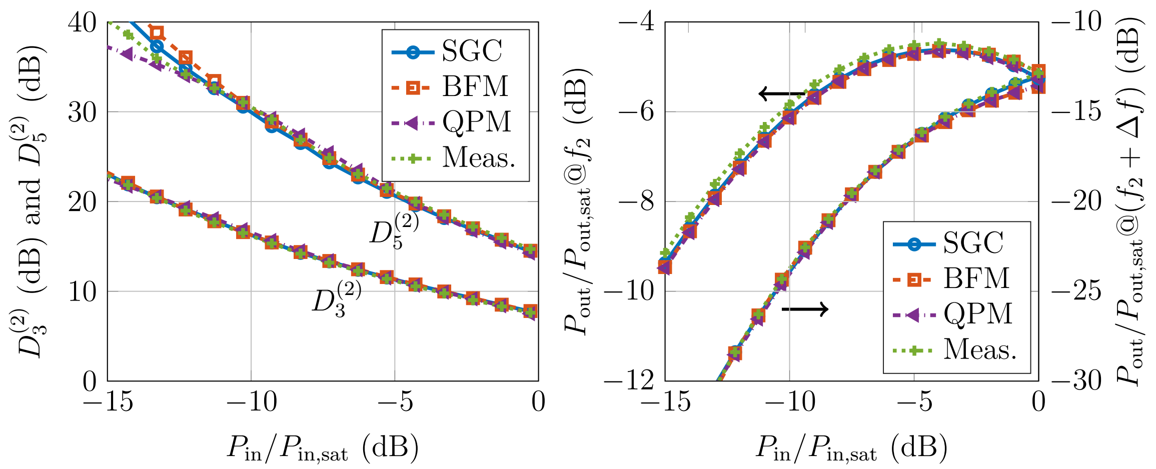
In these two parts, both the MVTRAD- and the measurement-based approach are evaluated with regards to two-tone excitation. Although TWTs are typically not excited with simple two-tone signals, these can easily be understood and are routinely measured by manufacturers. They represent the least challenging multi-frequency scenario and yield meaningful figures of merit. There are even approximate formulas readily available from literature, as was discussed in Section 3.3.1. Dual-tone signals are thus best suited to verify the intermodulation performance predicted by the approaches discussed here.

Afterwards, the broadband behavior is considered by means of communication signals, transient steps and noise-like signals. Also, the models are compared for a FW-TWT, which has a strongly dispersive behavior at a much higher frequency. In the end of the section, a short conclusion summarizes the results.

3.4.1 Multi-Tone Simulation with Measurement-Fitted Models

For the following comparisons, measurements are carried out on the reference communication Ku-Band TWT. The model parameters of the Envelope codes are extracted from CW AM-AM and AM-PM measurements throughout the frequency band of interest. In addition to the model base data, reference two-tone measurements with different frequency spacings are carried out, such that a direct comparison between Envelope simulation and measurement results is possible.

Therefore, a symmetric excitation is considered. The equal input power of the two tones is simultaneously altered from small-signal excitation to saturation in steps of 2 dB. The simulated and measured intermodulation products for a constant central frequency f_c and a frequency spacing of $\Delta f = 100\,\text{MHz}$ are shown in Figure 3.20a. As the lower side-band is nearly indistinguishable from the upper side-band, it is not shown. The intermodulation product ratio definition from Equation (3.8) is used. Both D_3 and D_5 show near perfect agreement. This can be explained by the slowly changing amplitude of the input signal and the almost constant amplifier characteristics in the relevant band for this scenario.



(a) Third- and fifth-order intermodulation product ratio at $\Delta f = 100$ MHz spacing.

(b) Main tone and intermodulation product at $\Delta f = 500$ MHz spacing.

Figure 3.20: Output power at main- and side-tone and intermodulation ratios for various measurement-based models compared to measurements.

The agreement still holds for a large frequency offset of $\Delta f = 500\,\mathrm{MHz}$. The SGC model shows a performance similar to the more involved methods, as can be seen in Figure 3.20b. Here, the power levels at f_2 and $f_2 + \Delta f$ are reported. While the overall agreement again is good, in this case all models underestimate the power at one of the excitation frequencies by about 0.15 dB to 0.18 dB. This discrepancy might stem from a ripple on the gain over frequency [62]. Such a ripple is typically mismatch-induced, i.e., due to imperfect input and output couplers, and thus is not considered in the models.

For communication systems, the phase distortion is likewise crucial. Therefore, with the phase transfer factor k_{T} as defined in Equation (3.21), a further metric is investigated. It is shown for different Δf versus input power in Figure 3.21. The phase transfer factor is obtained with an asymmetric excitation, where $P_{\mathrm{in}}(f_1) = P_{\mathrm{in}}(f_2) - 15\,\mathrm{dB}$. Then the phase difference at f_1 is measured, while reducing the power at the other tone $P_{\mathrm{in}}(f_2)$ by 1 dB. For a better overview, the absolute error Δk_{T} between simulation and measurements is also shown. As the nonlinear phase is strongly dispersive, the frequency dependent models yield improved predictions, as is discussed in the following.

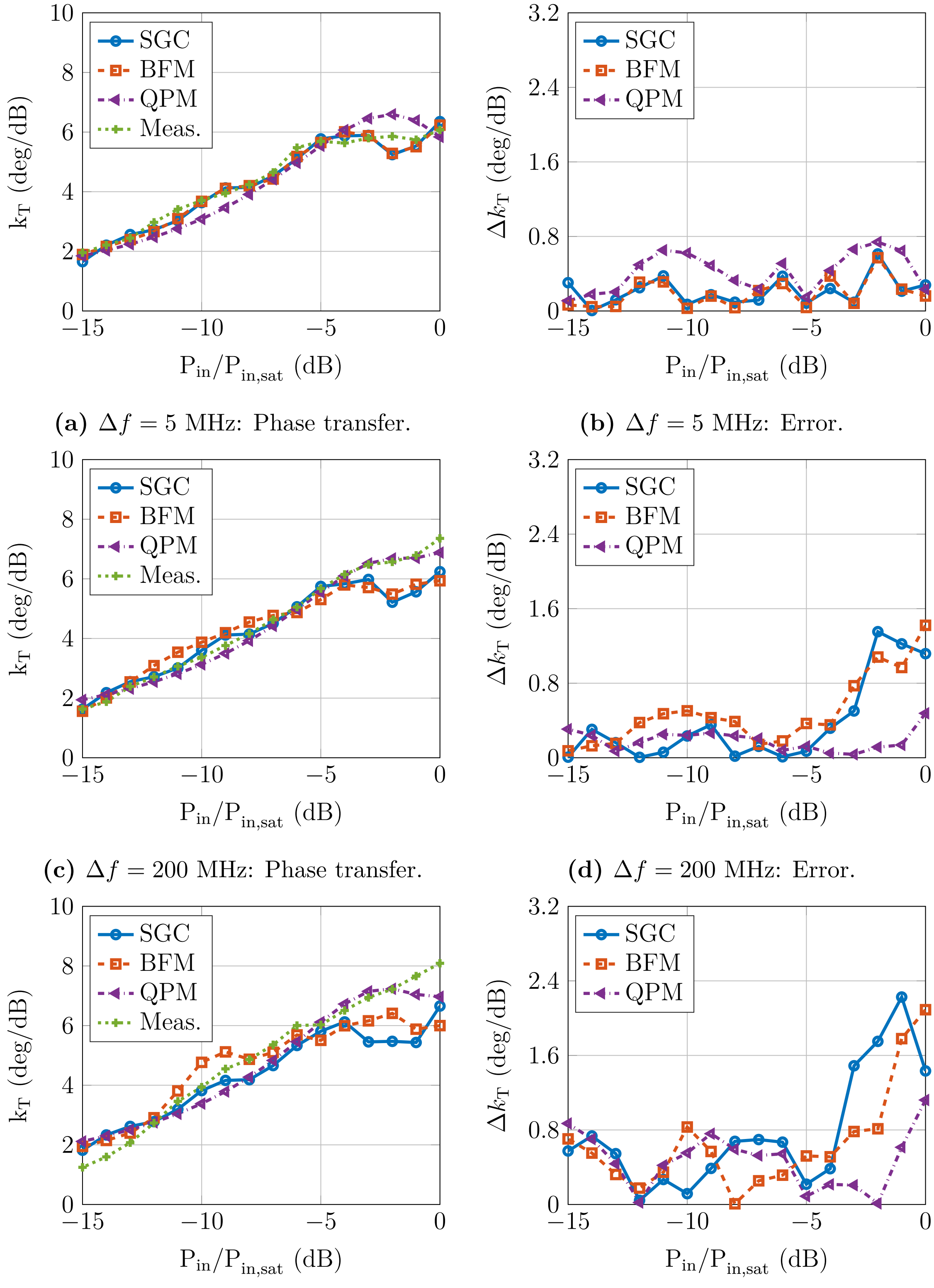
For two narrowly spaced tones at $\Delta f = 5\,\mathrm{MHz}$, shown in Figures 3.21a and 3.21b, all models accurately predict the measured results. In Figures 3.21c to 3.21f the advantage of adding frequency information to the model is illustrated. This is especially obvious near saturation where the phase error rapidly increases. While the error rises above 1°/dB for both SGC and BFM in the case of $\Delta f = 200\,\text{MHz}$, it stays below 0.5°/dB for QPM. Likewise, for $\Delta f = 1\,\text{GHz}$ the error is significantly larger for SGC and BFM, exceeding 2.1°/dB, compared to the QPM, where it remains below 1.2°/dB. Thus, only the QPM considerably benefits from the inclusion of dispersion information, although the same information is present in the BFM. This is due to the strong fluctuation and large range of the Bessel function fitting parameters over frequency, which yields a highly sensitive model. The QPM parameters, in contrast, change far more gracefully, such that, for instance, the required interpolation between fitting points is significantly improved.

3.4.2 Multi-Tone Simulation with MVTRAD-Fitted Models

In the preceding section, only the Envelope methods have been evaluated. To extend the study to the hybrid approach, an MVTRAD reference model has been generated and fitted closely to the measured TWT. The measured results are adjusted by the difference in saturated input power between the MVTRAD model and the measured data to enable comparison with multi-tone measurements. In addition, the underlying MVTRAD model is not only used inside the full simulation chain, but also for standalone frequency-domain simulation with multi-tone signals in MVTRAD and CST following Section 3.1.2 and Section 3.2, respectively. In the following, the legend entry MVT denotes MVTRAD simulation results.

As before, different dual-tone signals with the TWT driven close to saturation are measured and calculated with both the hybrid methods and MVTRAD alone. Therefor, CW output power and absolute phase versus input power and frequency are simulated in MVTRAD beforehand, to extract the parameters of the hybrid methods. The input power is increased from small-signal excitation to 10 dB into overdrive in steps of 1 dB and the frequency information is obtained in steps of 100 MHz throughout the band of interest.

For Figure 3.22a and Figure 3.22b, the center frequency of the considered signals is placed at the center of the amplification band, where the dispersion is naturally low by design. In Figure 3.22a the power spectrum for a two-tone signal with frequency spacing $\Delta f = 100\,\text{MHz}$ at 0.5 dB input back-off is shown. All models closely match both the measurements and the MVTRAD simulation results due to the small spacing between the tones. Still, at a distance of $3\Delta f$ from the frequency center, a small error can be seen, which is mainly caused by inaccuracies in the reference MVTRAD model.



(a) $\Delta f = 5$ MHz: Phase transfer.

(b) $\Delta f = 5$ MHz: Error.

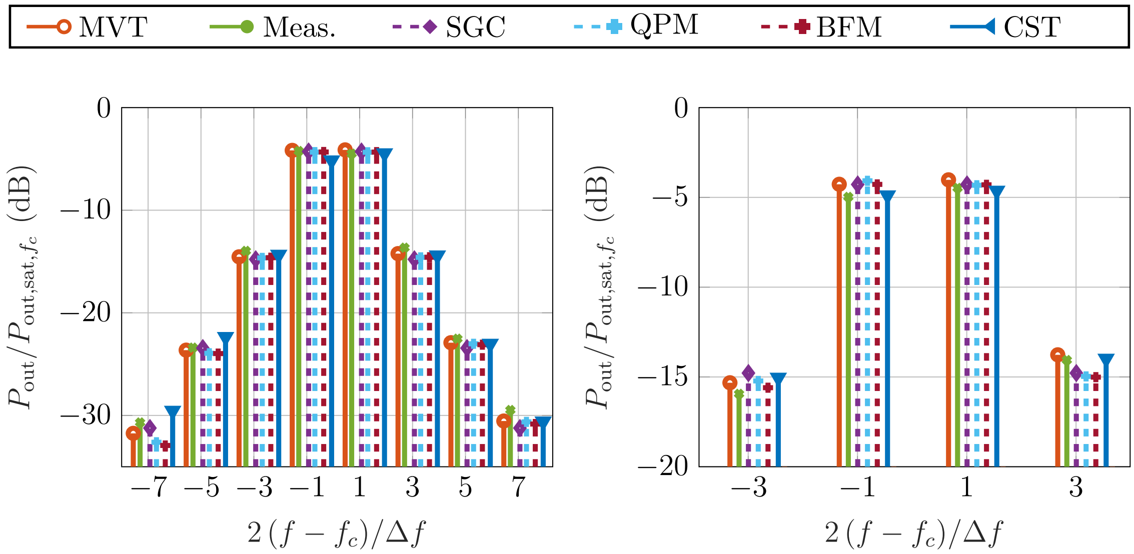
(c) $\Delta f = 200$ MHz: Phase transfer.

(d) $\Delta f = 200$ MHz: Error.

(e) $\Delta f = 1$ GHz: Phase transfer.

(f) $\Delta f = 1$ GHz: Error.

Figure 3.21: Phase transfer factor for different frequency spacings Δf.



(a) $\Delta f = 100$ MHz.

(b) $\Delta f = 500$ MHz.

Figure 3.22: Two-tone signals around f_c with a frequency spacing of Δf at 0.5 dB input back-off. The legend on top is valid for both plots.

The results for a larger frequency spacing of $\Delta f = 500$ MHz are shown in Figure 3.22b. The output power at the upper main frequency $P_{\mathrm{out}}(f_2)$ is slightly overestimated by up to 0.2 dB not only by the hybrid models but also by MVTRAD. The error is even larger at the other main tone with $P_{\mathrm{out}}(f_1)$ differing by as much as 0.5 dB. This deviation is not visible for the CST results, which hints at a model reference issue. Still, the frequency-dependence helps improving the estimation, as can be seen at $f_1 - \Delta f$. With the QPM the error shrinks to less than 0.5 dB compared to the SGC's overestimation by 1.2 dB. No further improvement is possible as the error is in a similar range in MVTRAD. For higher intermodulation products, all models show similar deviations.

In addition to the signals at the band center, two more dispersive scenarios are investigated in the following. For the results in Figure 3.23a, the center frequency is shifted down by around 5 % and a large frequency spacing of $\Delta f = 500$ MHz is chosen. At the two main-tone frequencies, all models perform similarly well compared to the measurements. On the low-frequency side of the output spectrum no reliable measurements are available. This is due to the input and output coupler, and limited calibration. Still, a comparison between Envelope methods, CST, and MVTRAD is attempted. Figure 3.23a shows that the resulting power spectrum from MVTRAD is slightly more asymmetric than in the previous cases, as the device is more dispersive. This is confirmed by the CST simulation. Contrary to this, due to its underlying mechanism, the SGC model always results in a symmetric output spectrum for a symmetric dual-tone input signal. Thus, it yields a large error of up to 1.5 dB. Both QPM and BFM are significantly closer. For the upper side-band, all hybrid models are similarly far away from either measurements, MVTRAD, or CST, with the error being around 1 dB to 1.5 dB. Finally, Figure 3.23b reports the results for a center frequency shifted further downwards to the lower band edge and a frequency spacing of $\Delta f = 200\,\mathrm{MHz}$. Again, both BFM and QPM yield some improvements at the side-bands, reducing the error at $f_2 + \Delta f$ to less than 0.3 dB compared to 0.7 dB for SGC. At $f_1 - \Delta f$ the results can only be compared to MVTRAD and CST. As before, the dispersive models significantly improve the estimation compared to the SGC, lowering the error from 1.2 dB to 0.5 dB.

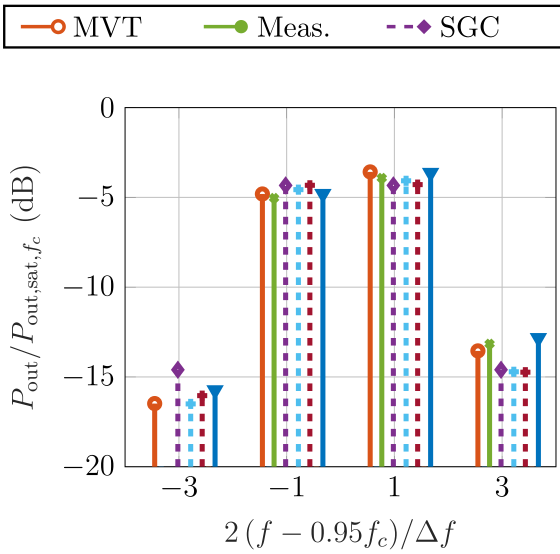
In total, one can observe a benefit of using the QPM or BFM concerning their applicability in a hybrid time- and frequency-domain TWT simulation. Still, another outcome of this investigation is that for a real communications TWT - as is investigated here -, at typical operating conditions and for practically relevant frequency spacings, the basic SGC model is sufficient to predict the relevant intermodulation power levels. On the other hand, when trying to predict the phase related properties, such as the phase transfer coefficients, the more sophisticated models present their advantages more clearly.

3.4.3 Transients and Communication Signals

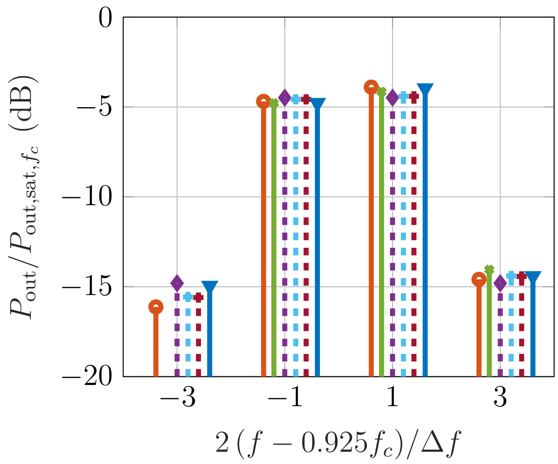
Sets of two-tone excitations, while they are popular for amplifier characterization, do of course not always provide a complete picture of TWTs in communication systems. Such an approach is strongly limited, as the initial phase relations between the two tones do not have an influence on the power spectrum. In addition, TWTs for communication satellites are operated with complexly modulated communication signals. Thus, communication system metrics such as constellation diagrams, eye-diagrams, and others are required to fully characterize the tube in realistic operating conditions.

Transients

The question of accurate frequency or phase representation is tightly connected to the one of transients and how well the models follow signal transitions from one state to another. Therefore, it is useful to consider various scenarios, where the signal is stepped up from zero-drive with different signal bandwidths.



(a) Tones around $0.95f_c$ with $\Delta f = 500$ MHz.



(b) Tones around $0.925f_c$ with $\Delta f = 200$ MHz.

Figure 3.23: Two-tone signals with a frequency spacing of Δf at 0.5 dB input back-off. The legend on top is valid for both plots.

In CST, arbitrary signal excitation can be defined in time-domain, e.g., by inserting it from an ASCII-coded input file. As the underlying model has a sound physical full-wave basis, the results serve as a benchmark for the Envelope methods. In particular, one can assume that a physical device cannot immediately follow a sharp ideal step in the oscillation, but smoothens it according to its physical bandwidth. This is mainly a result of the dispersive properties of the line.

In Figure 3.24 the output signal envelopes for several transitions simulated in CST, SGC and QPM are shown. For all plots, the curves are shifted in time, as the absolute delay is not taken into account and cannot accurately be matched between CST and the hybrid approach. Here, an absolute shift enables a visual comparison without corrupting the message.

Figure 3.24a reports the response for an ideal step, which is not filtered and therefore inhibits an infinite spectrum. It directly drives the models from zero-drive to a large-signal CW excitation. The physics-based simulation in CST shows a flattened response compared to the input, as expected. The observed behavior from SGC and QPM shows a major difference. The SGC model, not having any frequency-domain information available, directly follows the input as a scaled version of it. In contrast, the QPM model accurately reproduces the slope as predicted by CST. The waveform differs slightly, which is due to the limited spectrum considered in the QPM and the inherent harmonic representation problem of Envelope models.

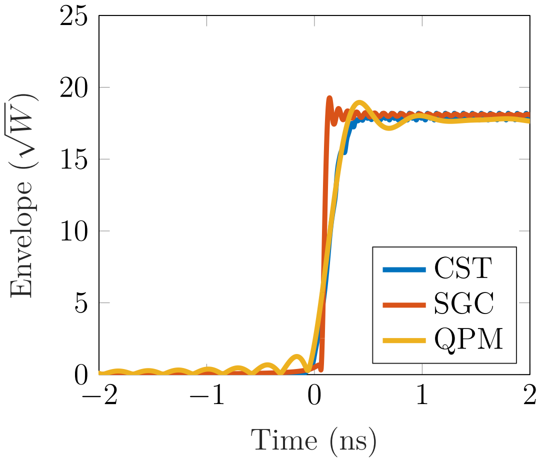
To investigate the frequency limits of the models, an RRC filter is used to limit the input signal to a finite bandwidth. Even for a large bandwidth of 1 GHz, the step is smooth enough for the SGC model to effortlessly follow it, which can be seen in Figure 3.24b. Also, the ripple in the waveform, as seen for the QPM results in Figure 3.24a, can be found in both CST and SGC for the filtered signal, which hints at the aforementioned bandwidth issue of the QPM with the ideal step response. In Figure 3.24c, two further transitions are considered. Both signals are RRC filtered with a bandwidth of 1 GHz. The step from zero-drive to small-signal drive shows that the sloping error is not a nonlinear effect but indeed caused by the dispersion of the line. The same can be seen for a step from a steady-state large-signal oscillation to a small-signal state. The small-signal level is not exactly reached, as the MVTRAD and CST models do not fit perfectly for both saturation and small-signal input. As mentioned in Section 3.4.2, the difference in saturated input power is compensated, leaving room for a deviation in the small-signal gain.

In conclusion, for realistic bandwidths in the considered Ku-Band communication scenario, even the SGC model accurately represents the transients. For higher bandwidths, where the SGC model is limited, the QPM is a viable alternative, behaving very similarly to the physics based full-wave model.

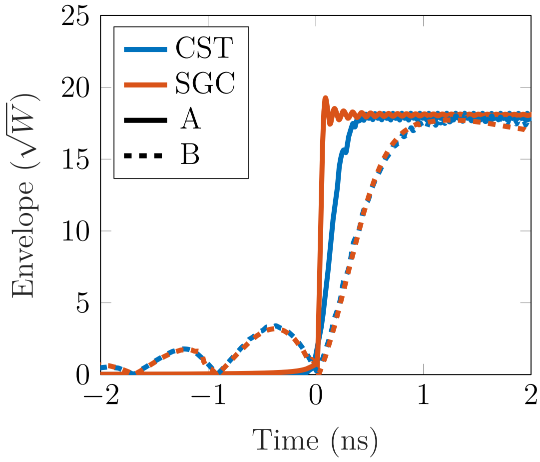
QPSK and Eye-Diagram

As discussed before, a full characterization of a TWT in realistic operating conditions requires communication signals, as for instance defined by the DVB-S2 standard. In Figures 3.25a and 3.25b a constellation diagram of a large-signal QPSK sequence from measurements is compared to simulation in SIMBA. As the signal bandwidth is chosen as typical for the Ku-Band, the SGC model suffices for the calculation, as was established in the previous sections. The roughly wedge-shaped point clouds at each constellation point are notched at one end. These apertures, which are not noise-related, but purely deterministic and a result of the different paths from one constellation point to another, are accurately predicted in simulation.

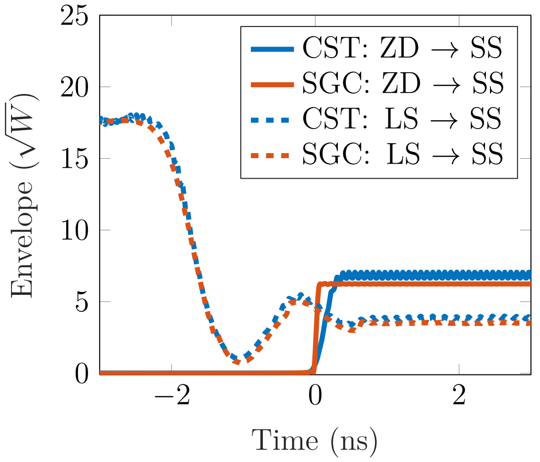
A similar accuracy can be seen for the higher-order modulation case in Figures 3.25c and 3.25d. Here, constellation diagrams of 16QAM signals are shown, again both for measurements and simulation in SIMBA. The drive level is 5 dB above saturation, strongly highlighting the nonlinear distortions.



(a) Unfiltered ideally stepped signal from zero-drive to large-signal excitation.

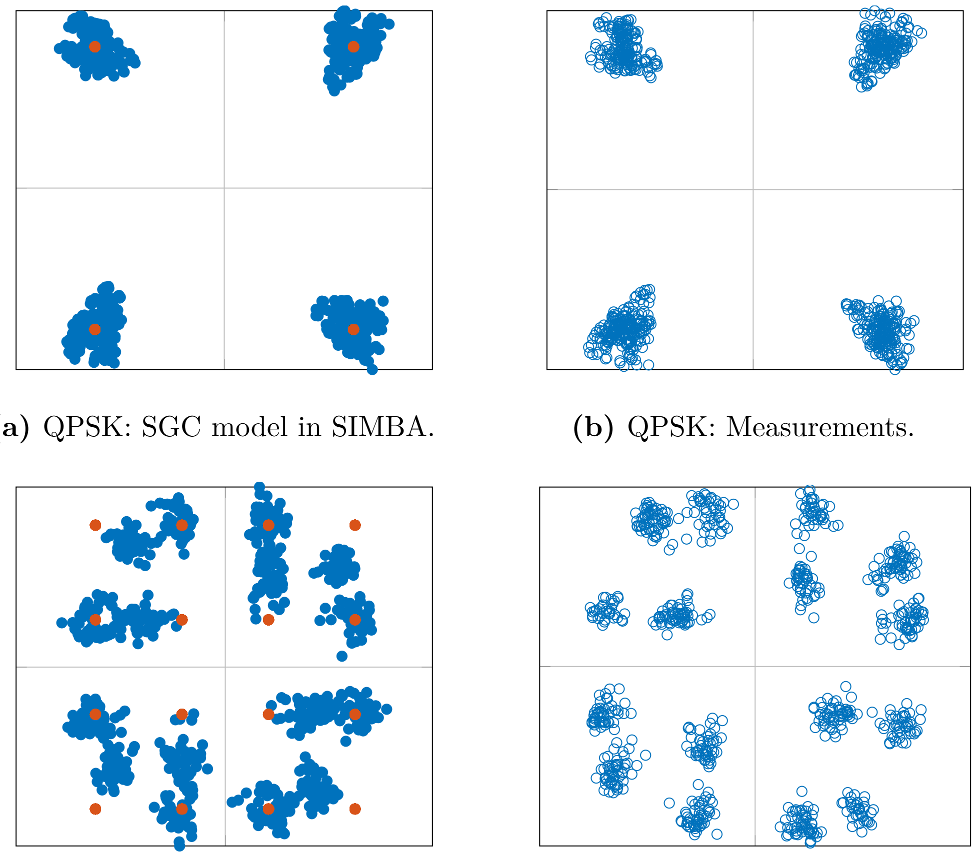


(b) Unfiltered ideal step (A), and 1 GHz RRC filtered step (B), from zero-drive to large-signal excitation.



(c) RRC filtered signal steps with 1 GHz bandwidth. (ZD): zero-drive, (SS): small-signal, (LS): large-signal excitation.

Figure 3.24: Transient processes in SGC and QPM compared to CST simulation.



(a) QPSK: SGC model in SIMBA.

(b) QPSK: Measurements.

(c) 16QAM: SGC model in SIMBA.

(d) 16QAM: Measurements.

Figure 3.25: Constellation diagrams from 800 symbols each, for a QPSK and a 16QAM signal at a drive level of 5 dB above CW saturation.

In Figure 3.26a the effective output power in the modulated case is shown in comparison to simulation. Not only do the results align well with the measurements, the behavior with higher-order modulation is also visible. A metric for the distortions occurring in a communication system is the so-called error-vector magnitude (EVM). It is defined as

$$\mathrm{EVM} = \frac{P_{\mathrm{error}}}{P_{\mathrm{reference}}} = \sqrt{\frac{\frac{1}{N}\sum_{k}\left[\left(I_k - \hat{I}_k\right)^2 + \left(Q_k - \hat{Q}_k\right)^2\right]}{\frac{1}{N}\sum_{k}\left[(I_k)^2 + (Q_k)^2\right]}}, \tag{3.58}$$

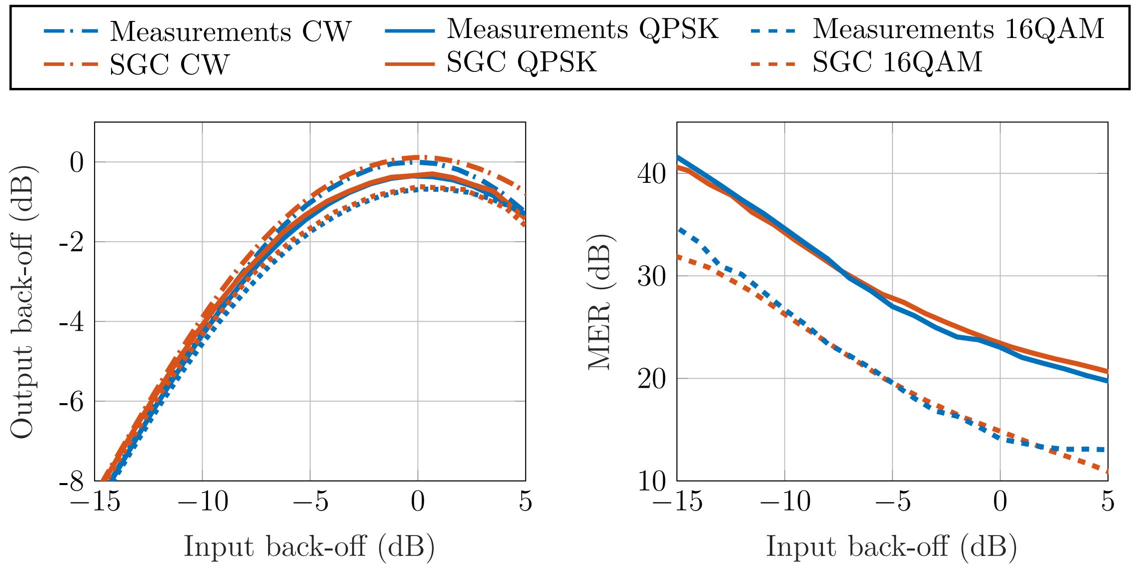
where $\hat{I}_k$ and $\hat{Q}_k$ are the measured inphase and quadrature component of the k-th symbol, I_k and Q_k are those of the corresponding reference point, and N is the number of considered symbols. When shown in logarithmic scale, it is sometimes referred to as the modulation-error ratio (MER) and calculated with

$$\mathrm{MER} = 10\log_{10}\left(\frac{P_{\mathrm{reference}}}{P_{\mathrm{error}}}\right). \tag{3.59}$$

In Figure 3.26b, the MER is shown for QPSK and 16QAM, comparing simulation and measurements. The agreement is good from saturation downwards to typical back-off values. For 16QAM, estimation in the overdrive region is particularly erroneous as the peak-to-average power-ratio (PAPR) is high compared to the QPSK case. Far in back-off, the deviations stem from the fact that the error vector there is small compared to the reference power, such that it is subject to stronger fluctuation.

Looking at the spectrum of a modulated QPSK signal, as shown in Figure 3.27, one can observe a fundamental limitation of the Envelope methods in the presented form, compared to CST. While the information on the higher harmonics is inherently connected to the considered transfer curves, components at these harmonics cannot be found in the output signal. Thus, while the in-band components of the modulated signal around a carrier frequency f_0 are correct, the corresponding intermodulation products around multiples of f_0 are not present. When information on the harmonics is required, physics-based simulation models, such as MVTRAD or CST, are therefore better suited. Still, for many cases no components other than those around the carrier f_0 are required, as the output coupler typically is not well-matched for these frequencies. Thus, the relatively low power generated there is reflected and absorbed in the sever.

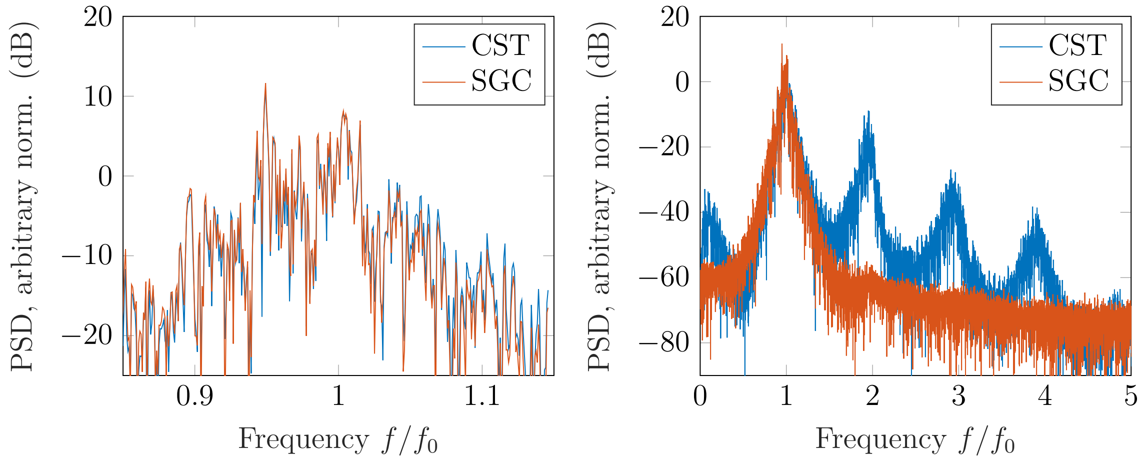
While both the Envelope models and CST's PIC solver can be excited with an arbitrary communication signal, this is more difficult for MVTRAD. Anyway, a full characterization with, e.g., an eye-diagram generation for a QPSK modulation requires an extensive Monte-Carlo study. While this is possible in measurements or time-domain



(a) Total output power.

(b) Distortions.

Figure 3.26: Output power and distortions of QPSK and 16QAM modulated signals from measurements and simulation. The legend on top is valid for both plots.



(a) Spectrum around the carrier frequency.

(b) Spectrum across several harmonics of the carrier frequency.

Figure 3.27: Output spectrum of a modulated signal with neighboring QPSK channel from CST and the SGC model.

simulation using Envelope codes, it would result in an infinite number of frequencies in a frequency-domain simulation in MVTRAD or a very large simulation time in a full-wave time-domain code as CST's PIC-solver. Thus, these two cannot be employed directly.

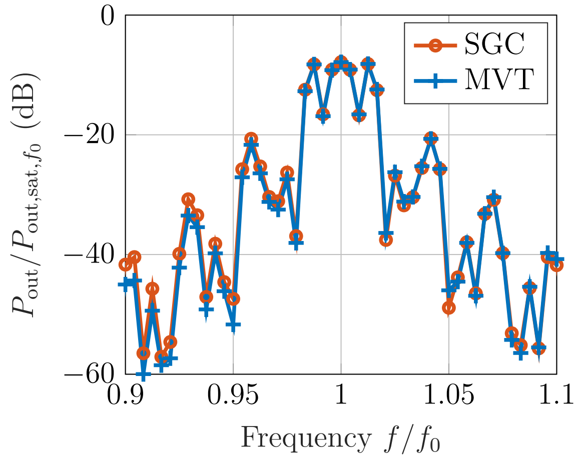
As the duration of a symbol is typically long, it is not feasable to simulate more than a few symbols in a full-wave code. Therefore, CST alone will only yield a limited picture. For MVTRAD, this is in principle even worse, as the continuous spectrum would require an infinite number of harmonics. Thus, a symbol reduction scheme has been conceived, limiting the number of simulation runs to a number of meaningful periodic ones. The reduction is based on a derivation in a noise-free, linear case, where the meaningful sequences are sufficient to depict the full eye-diagram. With nonlinearities, this is of course not the case anymore, but the chosen sequences still yield a good estimate of the general behavior. The reduction scheme and its underlying idea are outlined in [44]. The resulting periodic sequences can then be simulated in MVTRAD, and some comparisons to Envelope simulation for such reduced QPSK symbol sequences are presented in the following.

Therefor, the data rate R_c, the filter symbol span ξ, and the roll-off factor α of the pulse-forming filter are related to the fundamental frequency Δf, which is equal to the spacing between the discrete tones of the periodic signal, according to

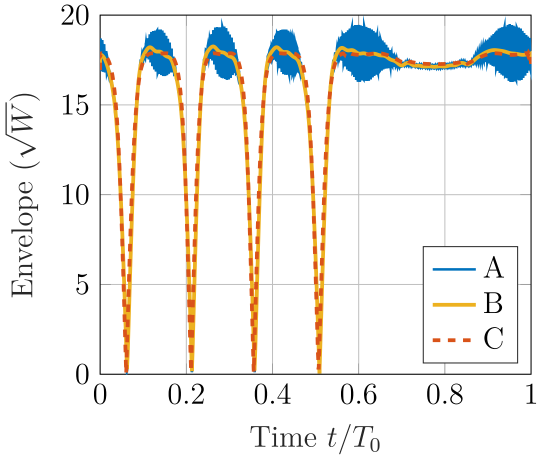
$$\Delta f = \frac{R_\mathrm{c}}{(1+\alpha)\cdot(\xi+1)}. \tag{3.60}$$

Here, Δf is chosen to 50 MHz. The carrier frequency is selected such that all discrete frequency components are multiples of Δf. Thus, the total frequency band containing significant information can be estimated to 350 MHz with a roll-off factor of 0.35. The generated signal is periodic with $T_0 = (\Delta f)^{-1}$.

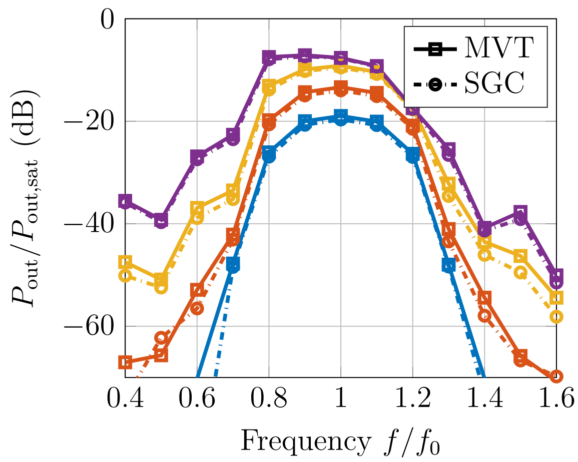
A complex phasor description of the Fourier coefficients at multiples of Δf representing the operating band and possible intermodulation products, as well as the corresponding second harmonic frequencies around $2f_0$ are computed and used as the excitation in MVTRAD. The computation time lies around four days per sequence as the number of harmonics and thus the number of required time steps within the base period is large. Then, using the phasor Fourier sum of the resulting amplitudes and phases, the resulting time-domain signal is reconstructed. After down-conversion and receive-filtering, the eye-diagram can be recalculated and used for simulation in the hybrid Envelope approach. In the following, an MVTRAD based SGC model is compared to a full steady-state simulation with MVTRAD. The operating point is chosen 2 dB below saturation to ensure strong intermodulation.



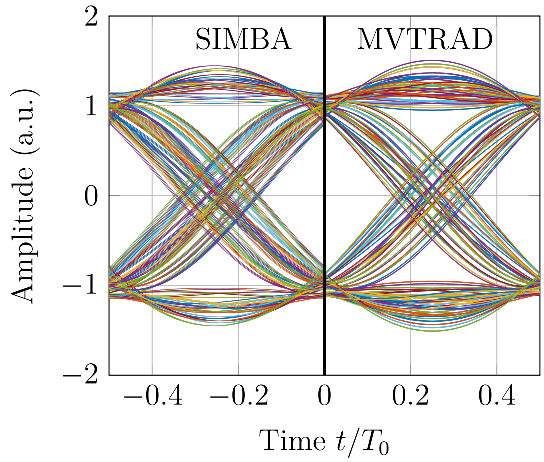
(a) Output power spectrum of an examplary QPSK sequence.



(b) (A): MVTRAD, with components around f_0 and $2f_0$. (B): MVTRAD, only components around f_0. (C): SGC model.



(c) Output power spectrum of another examplary QPSK sequence.



(d) Quadrature eye-diagrams from Envelope simulation and MVTRAD.

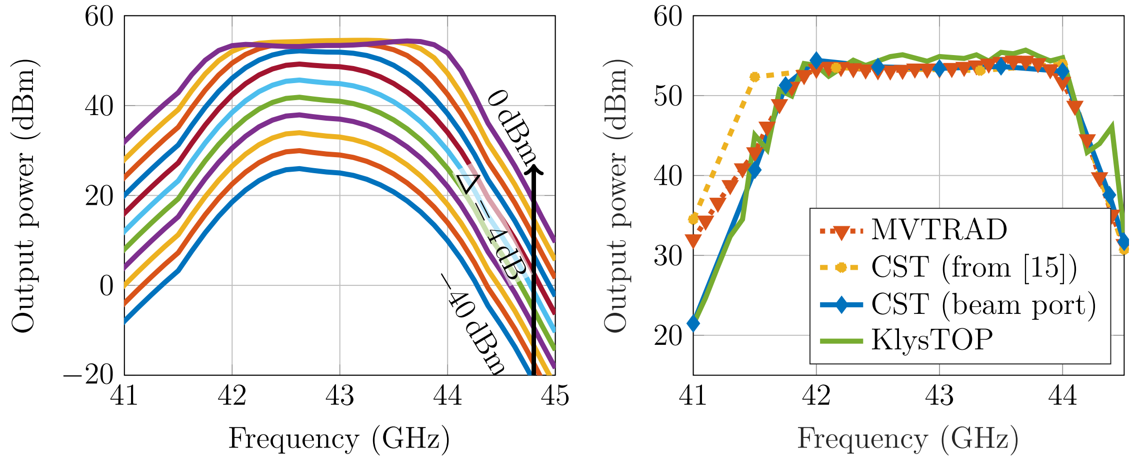
Figure 3.28: Exemplary periodic QPSK sequence simulated in MVTRAD and SIMBA.

Figure 3.28a and Figure 3.28b exemplarily show the power spectrum and the signal envelope of one period, generated using the Hilbert transform of the time-domain signals for one sequence, from both MVTRAD and the combined SGC-MVTRAD approach. Figure 3.28a reports an excellent overall agreement for the seven main input frequencies and still very good results for the intermodulation products. At the lower frequencies, the Envelope approach overestimates the power compared to MVTRAD. Following a conversion of the output power and phase of the individual frequency components from MVTRAD to a time-domain waveform, we see good agreement between the signals from SGC and MVTRAD, when only frequency components up to $f_0 + 40\Delta f$ are considered. Using the spectrum as calculated in MVTRAD, with components around $2f_0$, the modulus of the analytic signal representation looks slightly different. This is due to the distortions introduced by these second harmonic components, which are not seen from the Envelope approach. Anyway, as the signal is low-pass filtered on the demodulation side, they ultimately are irrelevant for the investigation. Figure 3.28c shows another sequence, highlighting that the spectral distribution of each sequence differs strongly. A comparison of the hybrid MVTRAD based SGC model with MVTRAD shows good agreement from small-signal excitation at 18 dB below saturation to large-signal excitation at saturation. It can be seen, that although the power spectrum is symmetric at the input, it can still be asymmetric at the output, as the phase relations of a signal containing more than two tones influence their respective power levels in a nonlinear device in an asymmetric manner.

Figure 3.28d finally compares the eye-diagrams obtained from MVTRAD and from the combined QPM-MVTRAD approach. For the Envelope results, both the vertical and the horizontal eye-opening are around 5 % smaller, as the eye is slightly more blurred. Still, the overall agreement between both diagrams is good.

3.4.4 Folded-Waveguide Simulation

The SGC and QPM have been thoroughly tested for typical communication helix TWTs. Here, there is no significant advantage of using the far more complex QPM approach compared to a simple SGC. With increasing frequency, starting around the Q-Band, more dispersive delay lines are needed, for instance the FW lines [15] as introduced in Section 2.1.2. The validity of the Envelope method, especially the QPM variant, is evaluated in the following. These FW-TWTs are strongly dispersive and possess a lower cut-off frequency which is to be considered. Thus, they constitute a more difficult problem, possibly reaching or illuminating the limits of the QPM.



(a) Output power over frequency for small- to large-signal input power levels, simulated in MVTRAD.

(b) Simulation results from CST, MVTRAD and KlysTOP at 0 dBm input power.

Figure 3.29: CW simulation results of the investigated FW-TWT.

In order to further investigate the model boundaries, a FW delay line structure from [15], [63] is investigated. Figure 3.29a shows the frequency-dependence and nonlinearity of the FW-TWT. Other quantities show similarly strong fluctuations over frequency, e.g., the saturated input power shows a difference of up to 25 dB across a 1 GHz band in the lower half of the investigated frequency range.

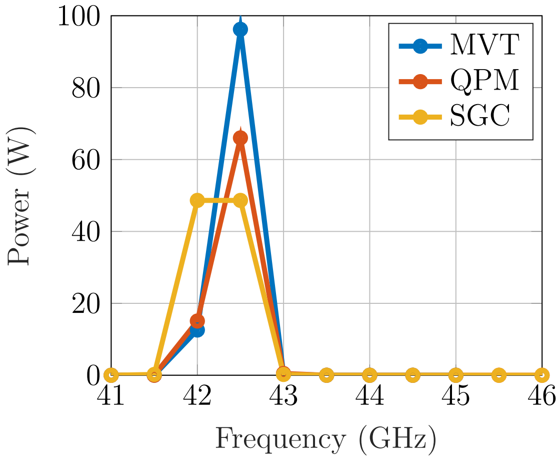
In [63], the chosen FW-TWT is described and simulated using MVTRAD, which is specialized for helical delay-lines. The FW's cold parameters have to be preprocessed to enable this. The simulation has been repeated using the specialized FW-TWT code KlysTOP [64], where a good overall agreement between the codes is found, hinting at the usability of MVTRAD for this task. In [15], simulation results obtained using CST are added. They exhibit a frequency shift of around 250 MHz compared to the other results.

With the same model as used for the CST results in [15], the beam injection is improved using the more involved particle injection procedure from Section 3.2.2. The beam setup file used in KlysTOP and MVTRAD is chosen as a reference for the beam interface in CST. As reported in Figure 3.29b, this corrects the frequency offset, such that the gain from the individual codes now agrees well.

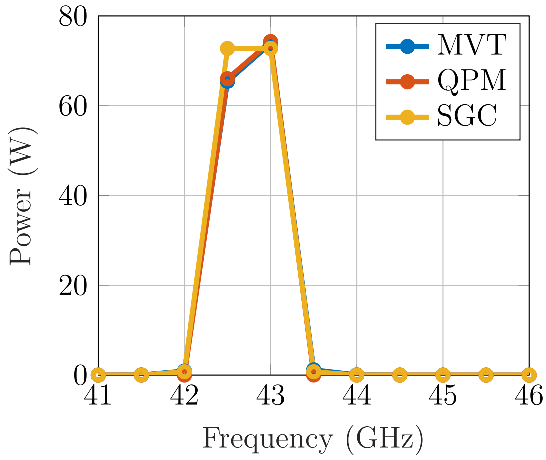
In order to extract suitable envelope models from the CST or MVTRAD data, a simulation of the saturation behavior at several frequencies within the amplification band is carried out. Subsequently a sufficiently fine-meshed simulation lattice is defined with respect to frequency and amplitude. For particularly dispersive structures, this means that a large number of individual simulation runs is required, since the respective saturation points are spread far apart. Also, it might be required for the amplitude range to include both small-signal and overdrive region for several frequencies. In addition, not only a high density over the frequency band is required, but also a wide range of frequencies is to be considered. In particular, the required distance to the lower cut-off around 39.5 GHz is unclear. There are therefore many, possibly lengthy simulation runs to be carried out. In the course of the investigations, CST therefore turned out to be unsuitable. Instead, the further investigations and the Envelope models are based on MVTRAD.

Two-tone excitations are considered to characterize the model performance. Different frequency spacings and center frequencies are selected in order to develop scenarios with varying difficulty. The range of possible frequency spacings is mainly limited by the model limits in MVTRAD. The calculation of the CW properties of the FW-TWT requires a higher spatial resolution of the numerical grid than for the helical ones. In addition, due to the narrow bandwidth of the structure, too large frequency distances do not make sense, but too small frequency spacings are not possible due to the high required numerical effort. Furthermore, the constraints given by the approach of the multi-tone calculation in MVTRAD with the many harmonics have to be considered. Therefore, in Figure 3.30, some scenarios from the amplification band edges to its center are investigated with frequency spacings of 0.5 GHz. It can be seen in all cases that the structure is too dispersive for accurate simulation with the SGC model. The drive frequencies already differ strongly, especially at the edges of the desired operating band. This is correctly predicted by the QPM. Still, at the lower end of the band, the results from MVTRAD and QPM differ from each other. It is not clear, whether any of the two is trustworthy there, as the lower cut-off frequency of the FW line is already close to the lower main tone.

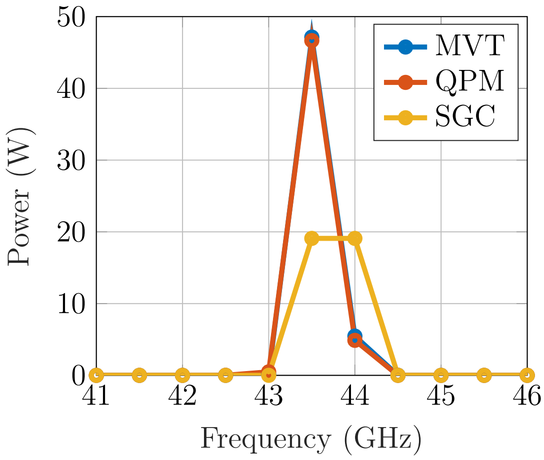
Figure 3.30f and Figure 3.30e highlight further differences between QPM and SGC model. In Figure 3.30e, the variation of the power ratio between the two main tones at the output from a symmetric two-tone excitation is plotted against the frequency spacing Δf, where the first tone is fixed at 43 GHz. In a small section around $\Delta f = 0$, the two output tones are close, such that the SGC model accurately predicts the output power ratio. Still, for a slightly higher spacing of $\Delta f = 0.4$ GHz the error already reaches 1 dB, increasing rapidly for higher Δf. The behavior is asymmetric, indicating that the maximum predictable signal bandwidth strongly depends on the position in the frequency band. In Figure 3.30f, the effect of a varying reference data extent (with respect to the considered frequency range) is investigated. Exemplarily, ranges of $\pm 2.5\Delta f$ and $\pm 4\Delta f$ around the center frequency are considered. The more



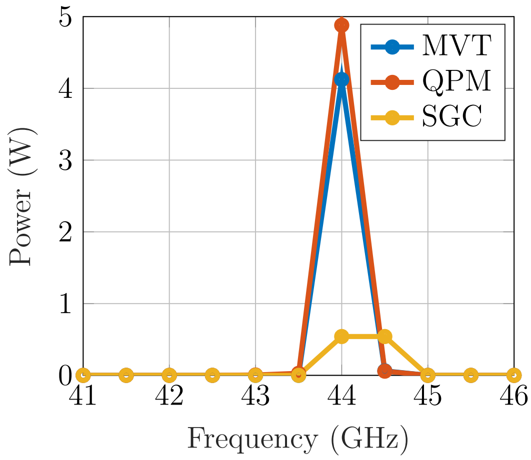
(a) Input tones at 42 and 42.5 GHz.



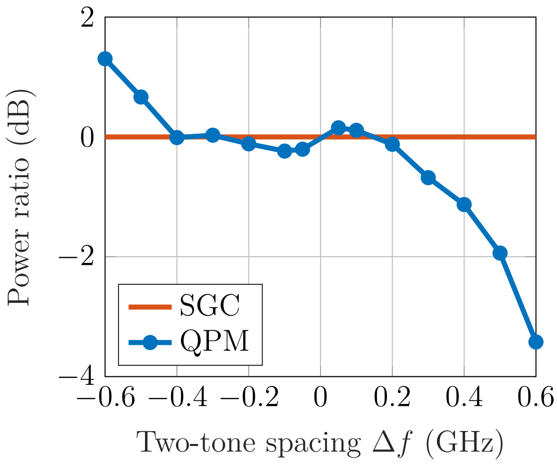
(b) Input tones at 42.5 and 43 GHz.



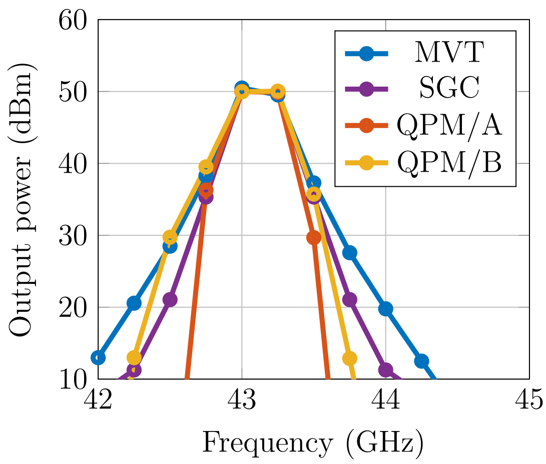
(c) Input tones at 43.5 and 44 GHz.



(d) Input tones at 44 and 44.5 GHz.



(e) Main-tone power ratio as a function of frequency spacing, with $f_1 = 43$ GHz.



(f) Two-tone results. (A): narrowband, (B): broadband reference data.

Figure 3.30: Simulation of dual-tone signals using MVTRAD and Envelope methods.

narrowband data range leads to a considerable deviation already in the neighboring intermodulation frequencies and overall leads to an erroneous estimation.

3.5 Conclusion

In this chapter, various possibilities of evaluating the multi-frequency behavior of TWTs have been assessed. The full-wave PIC approach in CST naturally enables the full characterization. Although modeling procedures to accelerate and simplify the model generation and simulation are presented, this still is a high-effort approach. It might be suitable for analysis purposes, with calculation times of around half a day for a CW excitation, but cannot be used efficiently at the design stage. In contrast, MVTRAD, which is already well-established for the delay-line design, can fulfill this task. Still, as the MVTRAD results are generated by calculating the greatest common divisor of the tones and scaling the time resolution according to the number of harmonics, this significantly increases the total interaction simulation time. Depending on the common fundamental frequency, the latter may exceed half a day for the presented scenarios. In principle, the MVTRAD multi-tone approach does not scale graciously, limiting the applicability for scenarios with many harmonics. For the Ku-Band TWT simulation, a comparison of the computation time in MVTRAD and in CST's PIC code for a number of two-tone excitations with varying frequency-spacing is listed in Table 3.2. It can be seen that while large frequency spacings can easily be simulated in MVTRAD and CST, the computation time rapidly rises, reducing their usability. They are especially not helpful for modulated signals with continuous spectra, except in some special, selected cases, as described in Section 3.4.3.

As an alternative, Envelope methods have been evaluated. The Envelope simulation itself can be done within minutes, the computation time for the hybrid approach is therefore mainly determined by the frequency-domain module. Employing parallel computing and considering as few frequency points as possible would reduce the total computation time to a few additional minutes. Two Envelope models have been evaluated in detail. For typical communication TWTs with helical delay-lines, e.g., in the Ku-Band, the basic SGC model has proven most practical, except for more involved phase-related properties. Here, the more intricate QPM is at advantage, although it requires more input data and thus in total the simulation takes considerably longer than with the SGC model. Also, for modern applications in higher frequency bands, for instance Q-Band, the QPM can surely play to its strengths, as more dispersive topologies are used.

Table 3.2: Comparison of approximate computation times in CST and MVTRAD. In the *fast mode*, only frequencies in the range of $(f_i \pm 5\Delta f)$ and second harmonics are considered.

Δf (MHz)	Period (ns)	MVTRAD			CST PIC
		Harmonics	Normal mode	Fast mode	(with GPU)
– (CW)	≈0.1	3	≈ 10 s	≈ 10 s	12 h
500	2	60	2.5 h	20 min	16 h
250	4	120	11.5 h	50 min	18 h
125	8	240	≫ 1 d	2.5 h	24 h
62.5	16	480	–	13 h	35 h
40	25	750	–	≫ 1 d	2.5 d
31.25	32	960	–	–	3 d

Thus, the usability of the presented simulation tools depends on the considered scenario. For continuous spectra and narrow bandwidths, the SGC model is the method of choice. For a broader bandwidth, the QPM is more accurate. Both the QPM and the SGC model are less helpful, the larger the tone distance is, as the computation time in MVTRAD and CST reduces to more practical ranges and their results are more trustworthy for larger frequency bands, due to their sound physical background. Of course, CST has the largest range of applicability, but as it is the most effort-expensive option, it mainly serves as a fall-back solution. Anyway, the simulation of random complexly modulated signals with very large bandwidths is not feasible in any of the tools, but also represents the least likely problem.

4 Back-Off Analysis and Optimization

In Chapter 3, different tools to evaluate the behavior of TWTs in modulated operation have been investigated. While the analysis of an already designed TWT is in itself an interesting purpose for these tools, they obviously can also be used to improve the device during the design process. This allows for the development of TWTs, which are especially suited for modern applications where increasingly complex modulation schemes are utilized. These are in principle less robust against distortions resulting from the nonlinear nature of the TWT. Therefore, these distortions are required to be as low as possible, which is ensured by driving the TWT far in back-off. Unfortunately, the high linearity comes at the cost of low output power and, which is even worse, a strongly diminished total efficiency. Comparing the total efficiency to the output power, one can see that the drive level at their maximum values, as well as their characteristics over drive, are slightly different. Therefore, in principle, a dedicated, separate optimization of the two quantities is imaginable.

There are two evident set-screws for the optimization of the TWT. As the main contributor of the total efficiency in back-off is the collector, a change in the collector will surely lead to a different efficiency reduction in back-off. It is advantageous to change the collector, since it is located at the end of the TWT, after the interaction has ended. This means that no change in the other amplifier characteristics is to be expected, but also that only little leeway is available. In contrast, the performance of TWTs strongly relies on the pitch profile of the delay line, as has been established in the previous chapters. Therefore, the design of this profile is essential for the adaptation of the TWT to its planned application, offering a large range of possible improvements of certain characteristics. Unfortunately, this usually happens at the cost of one or more other, possibly similarly important characteristics.

Within the framework of the validation of the different tools in Chapter 3, data was collected continuously. It has been shown that the dependencies between the design of the TWT itself, the usual gain characteristics, and the considered effective quantities in modulated operation cannot be read directly from the simulated and measured results. Instead, an initially quite complex, multidimensional problem with many possible set-screws is revealed. Their influence cannot be clearly described due to manifold interactions. Thus, it is difficult to make quantitative statements about how, for example, the delay line would have to be adapted in order to carry out an optimization in the back-off range for common and future modulation types.

4.1 Analysis of Nonlinear Effects

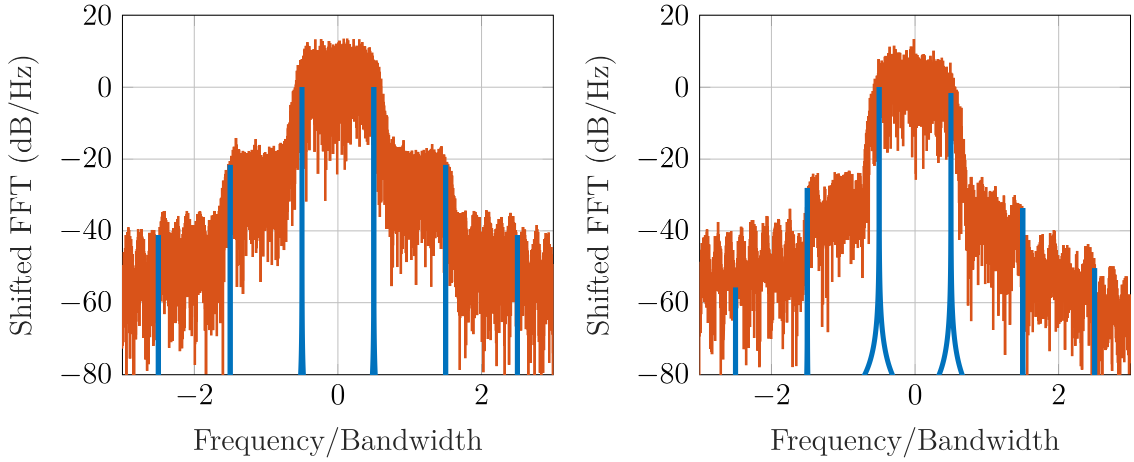
In the following, at first the impact of the nonlinearities on typical amplifier characteristics is evaluated, to enhance the understanding of the relationship between classical and effective modulated characteristics. Then, with these results in mind, ways to tackle the problem of back-off optimization are investigated.

4.1.1 Asymmetric Intermodulation

Simulation and measurements of dual-tone intermodulation product ratios have shown little variation over frequency in relevant bandwidths for typical communication TWTs. Still, for large frequency spacings Δf between the tones, a more slanted power spectrum is found, as expected, as the signal then spans over frequencies with significantly different transfer properties. The analysis of this slanting is rather complicated when considering communication signals, as in principle it depends on the actual modulation scheme. Also, the tilting of the spectral properties due to amplitude and phase distortions cannot be separated and figures of merit can hardly be defined unambiguously. Thus, as before, two-tone signals are taken into account.

Figure 4.1 shows QPM simulation results for a dispersive and a non-dispersive scenario, comparing two-tone excitations to QPSK modulated signals at equal input powers. For the sake of visual comparability, the power spectral density obtained from the output waveforms by means of an FFT is shifted, such that the two-tone peaks at the input frequencies are positioned at the average in-band value of the modulated signal spectrum. It can be seen that the dual-tone response serves as a reasonable indicator for the communication signal's spectral behavior and the expected slanting. Further experience gained throughout the verification study of the models in Chapter 3 leads to a similar conclusion. Therefore, the well-defined two-tone figures of merit are considered to evaluate the spectral tilting.

In Figure 4.2, the intermodulation product ratio as simulated with MVTRAD is investigated over frequency and drive, using the highly dispersive alternative reference Ku-Band TWT, highlighting the asymmetry between the powers at the upper and



(a) Non-dispersive case.

(b) Dispersive case.

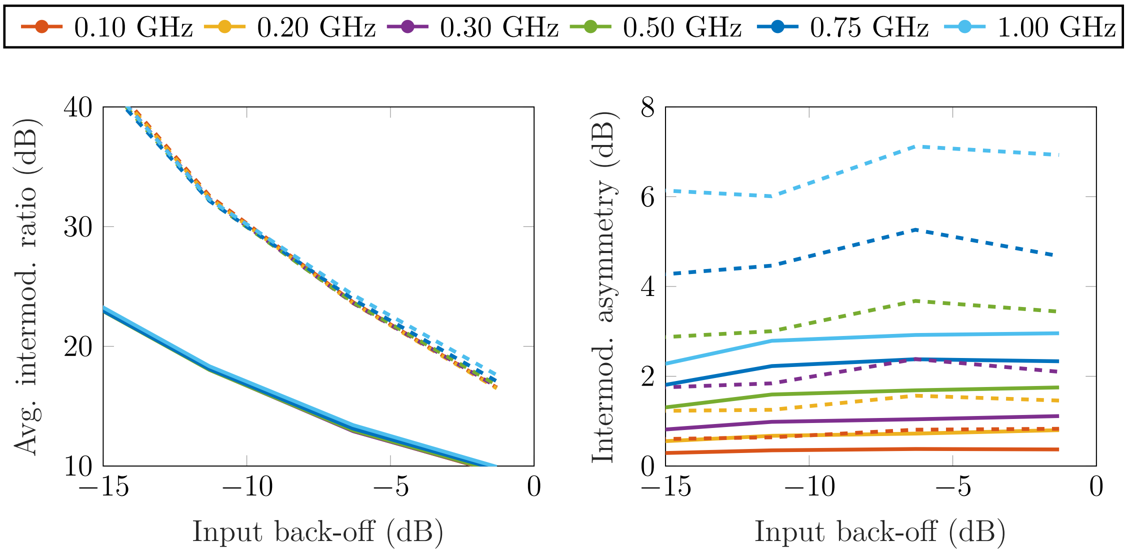
Figure 4.1: Simulation of equal-power two-tone and QPSK signals in SIMBA.

the lower intermodulation frequency. The lower tone is fixed at a frequency near the center of the band and the higher one is shifted from it by Δf.

The results for D_3 and D_5 show a similar behavior. While the average intermodulation ratio obviously decreases strongly towards saturation, the curves do not vary much with Δf. More surprisingly, the asymmetry, expressed here by the difference between upper and lower intermodulation ratio, is nearly constant over drive, especially in comparison to the variation of the intermodulation level with the input power. In contrast to Δf itself, which has a clearly visible impact, the degree of distortion seems to only marginally influence the asymmetry. This has indirectly been seen in the Envelope models, as this hints at the separability and chainability of frequency-dependence and nonlinearity, allowing the application of the proposed models.

Another interesting property is the increase of the asymmetry for higher order intermodulation products. This can be understood as a result of the larger impact of the observed gain differences for higher order products and the corresponding frequencies being further apart, typically resulting in a larger variation of the TWT's properties. Anyway, this finding is not of high relevance, as the intermodulation ratio rapidly increases with its order and the asymmetry occurs around a roughly constant average. Typically only the third order products are within a relevant power range in back-off.

A similar study is conducted for the phase transfer. It is asymmetrically defined, as one of the two tones is classically chosen significantly smaller than the other one. As typically only small bandwidths are considered, it is often not even specified which one

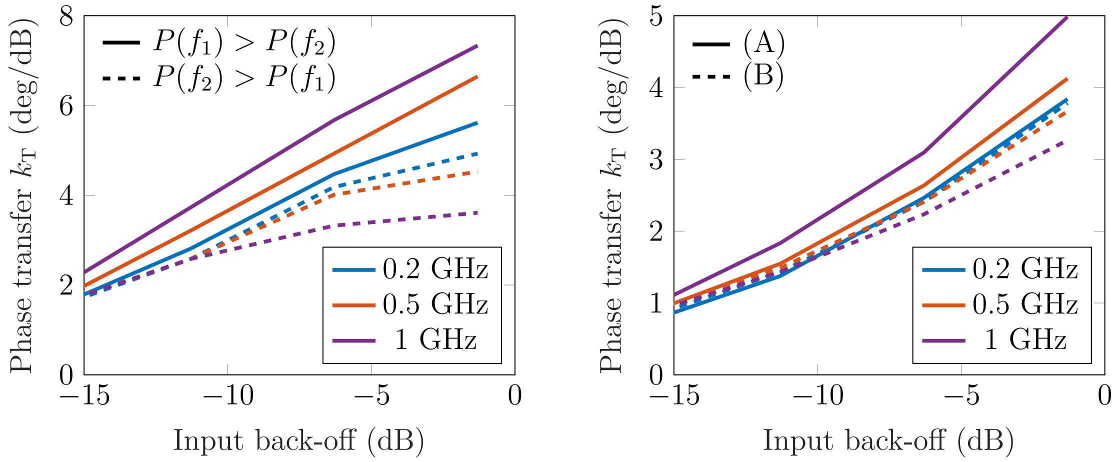


(a) Intermodulation product ratio. **(b)** Intermodulation product asymmetry.

Figure 4.2: Asymmetry of the intermodulation products over frequency and drive. The legend on top is valid for both plots and denotes the frequency spacing. The third order values are denoted by solid lines, the fifth order values by dashed lines.

of the carriers is supposed to be weaker. Here, considering three different frequency spacings from 0.2 GHz to 1 GHz, the asymmetry of k_{T} is investigated. For this purpose, using MVTRAD, first the lower tone is varied, recording the effect on the phase at the higher frequency, and afterwards the opposite problem is simulated. As for D_3 and D_5, the lower tone f_1 is fixed at a frequency near the center of the band and the higher one f_2 is placed at $f_1 + \Delta f$.

Figure 4.3a shows the phase transfer in back-off for both cases. The difference between the two increases with Δf and input drive level. Their mean value does not depend on Δf, though. This again hints at the separability of the functional blocks as mentioned in Section 3.3. A larger frequency spacing relates to a larger asymmetry, which differs depending on whether the tone at f_1 or the one at f_2 is altered. Additionally, it can be seen that the phase transfer is generally larger when the lower frequency is driving the TWT, as the phase change is observed at a higher frequency, and phase effects typically amplify in the direction of larger frequencies. Surprisingly, when operating at f_2, the phase transfer further in back-off does not seem to be a function of the spacing between the tones. This could be explained by a similar back-off level at f_2 for all three Δf.



(a) Unequal input drive with 15 dB difference between the two tones.

(b) Equal input drive. (A) Phase change at f_2 from drive change at f_1, and (B) vice-versa.

Figure 4.3: Asymmetry of the phase transfer k_T versus frequency spacing Δf and input drive for equal and unequal carriers, with $f_1 < f_2$. Input back-off here denotes the ratio between the total input power at f_1 and f_2 and saturation at f_1.

Instead of an asymmetric excitation, a phase transfer can also be defined for equally driven carriers. This is shown in Figure 4.3b, where the k_T values are in general lower than in Figure 4.3a. This complies with the theory in [37], where for low input powers a factor of two is estimated. Again, the spread depends on the frequency spacing, but it is more asymmetric now, as both tones drive the TWT. The mean value changes little, as in the unequal case. For instance, at around 10 dB in back-off, the average k_T from the equal carrier case lies between $1.4\,^\circ\,\mathrm{dB}^{-1}$ and $1.6\,^\circ\,\mathrm{dB}^{-1}$, and from the unequal one it lies between $2.7\,^\circ\,\mathrm{dB}^{-1}$ and $3.2\,^\circ\,\mathrm{dB}^{-1}$.

4.1.2 Impact on Communication Signals

As is seen, for instance, in Section 3.4.3, the interpretation of distortion metrics depends on the considered signal itself. Differently modulated signals have different underlying envelope distributions and encounter differently distributed distortions, even if their average power is similar. Figure 4.4 shows the variation of typical characteristics with various modulation schemes, i.e., from a simple QPSK to a complex 256QAM signal. The simulation is carried out in SIMBA, using a frame of 8000 symbols and a signal-to-noise ratio (SNR) of 20 dB.

Figure 4.4b shows the effective bit energy per noise power spectral density (EbNo). In contrast to the classical EbNo definition, where the bit energy is only related to the power density from an added noise source, the effective EbNo is calculated from the error vector due to the combination of added noise and distortions in the TWT:

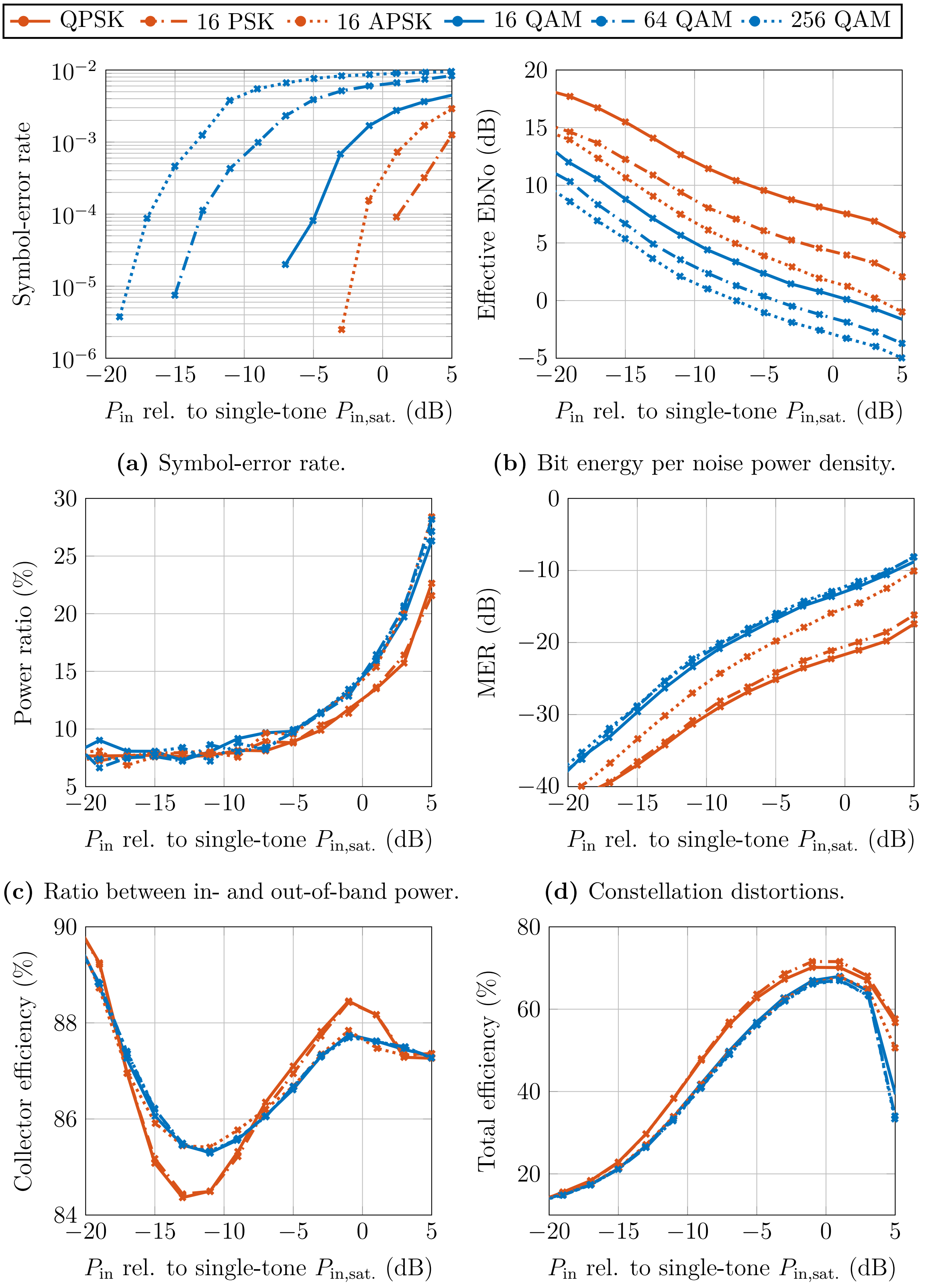
$$\mathrm{EbNo}_{\mathrm{eff.}} = 10\log_{10}\left(\frac{\frac{1}{N}\sum_k\left[(I_k)^2+(Q_k)^2\right]}{\frac{\log_2(M_\mathrm{s})}{N}\sum_k\left[\left(I_k-\hat{I}_k\right)^2+\left(Q_k-\hat{Q}_k\right)^2\right]}\right). \tag{4.1}$$

Here, M_s denotes the number of points in the constellation diagram. The effective EbNo additionally takes the signal distortions into account and therefore relates to an effective noise power. As expected, a high symbol-error rate (Figure 4.4a) corresponds to a low effective EbNo. This is in accordance with theory, from which it is known that higher order modulation schemes are more sensitive to noise, as the constellation points are effectively closer for equal average power.

Similarly to the intermodulation products from a two-tone excitation the distortions for a modulated carrier yield a staircase-like spectrum to both sides. Thus, undesired power is generated in neighboring bands. Larger distortions correspond to a larger share of power found outside the carrier bandwidth. Therefore, it is helpful to consider the ratio between in- and out-of-band power, as is shown in Figure 4.4c. For higher drive levels, one can see that the QPSK and 16PSK signals with their low PAPRs show significantly lower distortions than the other schemes. With increasing complexity, the distortions increase, but converge towards a limiting curve. The same holds for the MER, as shown in Figure 4.4d, which also increases for more complex modulation. Here, a difference between 16QAM and 16APSK is visible due to the different distribution of constellation points on two rings of equal power for 16APSK and three for 16QAM.

Finally, the effective efficiencies in Figure 4.4e and Figure 4.4f show that involved multi-amplitude modulations flatten the behavior over drive, as they consist of a broader range of amplitude levels over time, resulting in a weighted average of the CW efficiencies. The flattening does not occur boundlessly, but does not change much between the higher-order modulation schemes.

Typically, to reduce distortions at high output powers, TWTs are operated in combination with a linearizer. As the linearizer is inserted in front of the TWT, the now distorted signal envelope distribution at the input of the TWT might significantly change compared to a setting without linearization. A comparison of the envelope distributions from two exemplary modulation schemes, the 16APSK and the 64QAM, is presented in Figure 4.5. Figure 4.5a and Figure 4.5b show the envelope distribution for a 16APSK modulation at the input and output terminals of the TWT, both with and



(a) Symbol-error rate.

(b) Bit energy per noise power density.

(c) Ratio between in- and out-of-band power.

(d) Constellation distortions.

(e) Effective collector efficiency.

(f) Effective total efficiency.

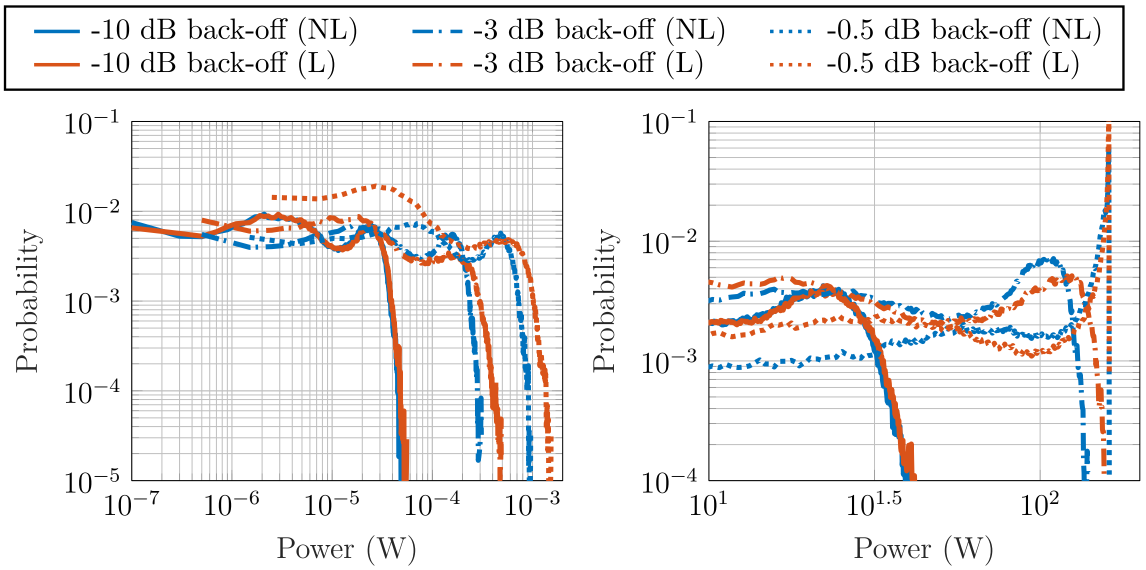
Figure 4.4: Nonlinear distortions and efficiency reduction over drive from SIMBA for a non-linearized Ku-Band TWT. The legend on top is valid for all plots.

without linearizer. To depict the behavior over drive, three different output back-off levels are considered, starting in small-signal regime at 10 dB below saturated output power, up to large-signal excitations at 3 dB and 0.5 dB below saturation. The same is shown in Figure 4.5c and Figure 4.5d for 64QAM modulation. In both cases, one can see that for small-signal excitation, as expected, no considerable changes in the input or output distribution are observable. As the gain compression and nonlinear phase shift are small, the linearizer does not need to alter the signal strongly. This can also be observed in the effective EVS plots in Figure 4.5e and Figure 4.5f. In contrast to this, for large-signal excitation, the envelope distribution shows a significant deviation in front of the TWT. Instead of an upwards power-shifted distribution, it is modified in a way that looks flattened and more spread. This increases closer to saturation. Fortunately, as the spreading only distributes the curves further, the effect on the EVS is small. As it is a result of averaging, the altered probabilities in parts cancel each other out. For a 3 dB output back-off point, effectively no change in the collector efficiency is to be expected between the linearized and the non-linearized setting.

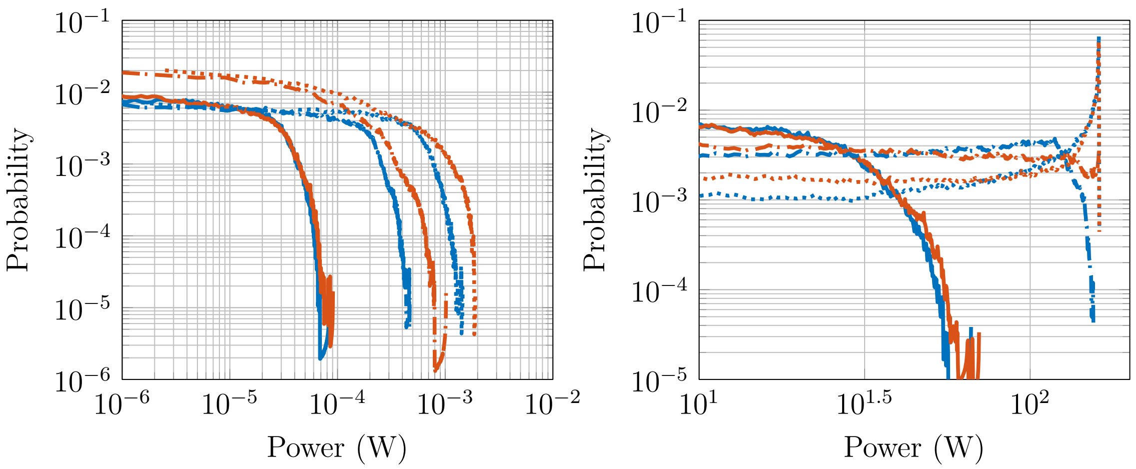
Obviously, other characteristics, especially the linearity, are altered more significantly. Thus, simply comparing the two settings at the same back-off operating point does not provide a complete picture. Figure 4.6a shows the effective EbNo with and without linearizer at different output power levels for different modulation schemes. A constant SNR at the input is assumed, resulting in a base EbNo-level for small-signal excitation, which depends on the size of the modulation alphabet[1]. As soon as the nonlinearity-based noise dominates the noise term, again roughly parallel lines can be observed, while the offset is much different from the one from small-signal excitation. With the linearizer, the same degree of distortion can be achieved at higher output power. Following this it might, for instance, be of interest which change in drive is required when changing from one modulation type to another and what this means for the EVS. This is shown both with and without linearizer in Figure 4.6b. Here, for each of the curves the output back-off value is chosen to reach an effective EbNo of 20 dB. The change in drive together with the predistorted signal distribution results in a significant change in the EVS and therefore not only in the beam efficiency, but also in the collector efficiency. This relates to a considerably different optimal collector stage voltage selection depending on the modulation scheme.

Thus, for the actual optimization of the TWT, it is beneficial to know whether it is supposed to be linearized or not, and which modulation is chosen. Following from that,

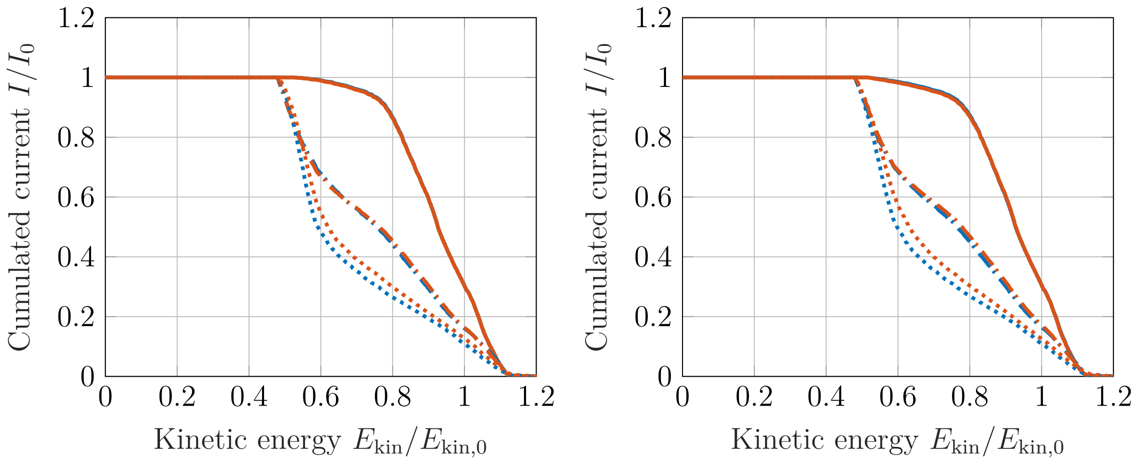
[1] As input SNR, bandwidth, and signal sampling are fixed, the ratio of symbol energy to the noise power spectral density is the same for all signals. In that case, the EbNo base-level variation only stems from the relationship between symbol energy and bit energy.



(a) 16APSK: Input envelope distribution. **(b)** 16APSK: Output envelope distribution.

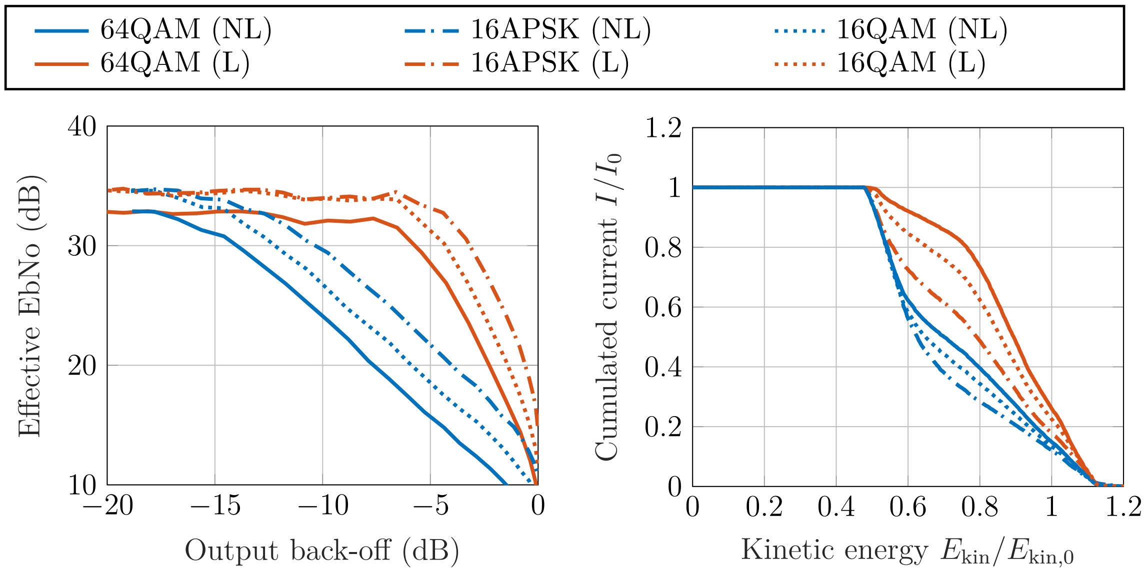


(c) 64QAM: Input envelope distribution. **(d)** 64QAM: Output envelope distribution.



(e) 16APSK: Effective EVS. **(f)** 64QAM: Effective EVS.

Figure 4.5: Simulation with (L) and without linearizer (NL). The legend on top shows output back-off values and is valid for all plots.

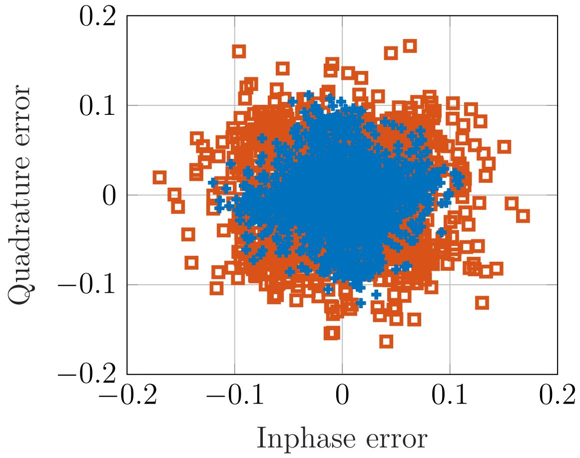


(a) Distortions versus output power. **(b)** EVS with and without linearizer.

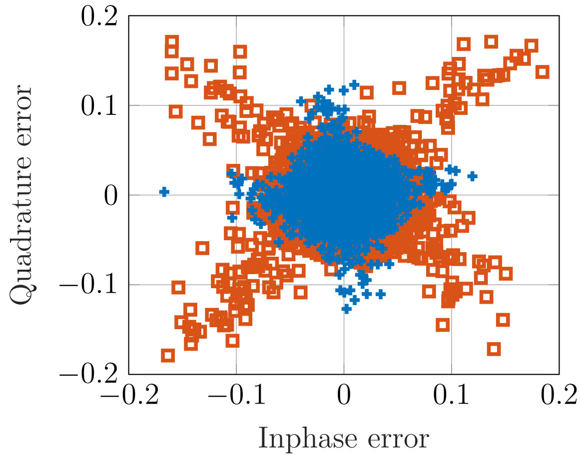
Figure 4.6: Equal-distortion based comparison with (L) and without linearizer (NL). The legend on top is valid for both plots.

the choice of the operating point plays a major role. For the operating point, various figures of merit can be taken into account. At maximum efficiency, typically near the saturated output power, too much distortion is introduced. Too far from saturation, distortion is not problematic anymore, but possibly power combining techniques, and thus more amplifiers, are required to reach the required signal power. Somewhere in between, one could expect some sort of sweet spot, where no characteristic is too bad. This spot consequently depends on the application.

It is also important to note that the effective EbNo or EVM, while it is a useful metric for the distortion, does not directly translate to a symbol-error rate, as typically presumed in literature. This is, as the theoretical relationship between EbNo and the symbol- or bit-error rate is usually calculated assuming distortions by an additive white Gaussian noise channel. Figure 4.7a and Figure 4.7b report the error-vector distribution with and without linearizer for which approximately the same effective EbNo is achieved. Not only is the distribution different from what would be expected from additive white Gaussian noise, the linearized and non-linearized error vectors are also different from each other. Both effects are especially strong for the 64QAM. Thus, while the effective MER will agree, the symbol-error rate might differ and is to be considered separately.



(a) 16APSK.



(b) 64QAM.

Figure 4.7: Equal-distortion based comparison of the error-vector distribution with (blue) and without (red) linearizer.

Unfortunately, the linearizer itself is generated based on the TWT, and the combined amplifier in reality is somewhat different from an ideal limiter. As a result, it is not so easy to consider a correctly linearized amplifier in an optimization of the TWT, but only to estimate the linearizer according to the TWT, for instance based on experience. Thus, in contrast to this analysis, where actual characteristics from a real TWT and the real, associated linearizer could be used, the optimization routine proposed in Section 4.3 only considers a non-linearized TWT.

4.2 Relating Single- and Multi-Tone Characteristics

To understand how to adapt the TWT's geometry in the most promising way towards its operation with complex modulation schemes, one could think of using a sufficiently large number of simulated and measured TWT data and a more or less intelligent pattern recognition, either human or computer-based. However, such an approach is not only highly uncertain with regards to the gain in knowledge, but also disproportionately time-consuming and resource-intensive, due to the required amount of data to be generated, as previous experience has shown.

Thus, to allow some degree of understanding, the development of suitable expressions to directly describe the relationships between - desirably - the structure, CW amplifier quantities, and communication characteristics is investigated in the following.

Classical approaches for partial problems known from literature (e.g., [17]) already show that this is not a simple task either. Even the connection between cold line, beam quantities, and classical amplifier characteristics can only be described analytically for very simple cases. For example, the amplification in the small-signal range for a homogeneous line is given in Equation (2.11). However, as soon as inhomogeneities and nonlinearities are to be taken into account, finding such expressions is rather challenging. Reasons for this can be found in the origin of the nonlinear effects from the ballistic behavior of the electron beam and the complex interaction between the coupled systems as described in Section 2.3. However, these difficulties are less grave in the development of expressions for the relationship between single-tone and multi-tone quantities. This is due to the fact that the physical causes and effects are already inherent components of the usual amplifier characteristics and the task can therefore be lifted from a purely physical to a model level. This is also the basis for the usability of Envelope based simulation tools as investigated in Section 3.3.

4.2.1 Low-Level Noise-Power Ratio Estimation

During the previous investigations, it was reported that different types of modulation have a different influence on common effective communication metrics. A characterization of the amplifier therefore always depends on the respective excitation. However, it has also been shown that, for a fixed TWT design, each of these metrics tends towards a limit when increasing the complexity of modulation. This happens as the signals then become more and more similar to noise from a spectral point of view. Therefore, it makes sense to use such a "noise-like" excitation as a means of characterization. Likewise approaches have been established for a long time. Therefor, a band-limited, notch-filtered input signal on the TWT is used to calculate the NPR, which has shown to be useful to evaluate the distortion performance of the amplifier [65]–[67]. At the output, the nonlinearity of the TWT introduces many additional frequency components, e.g., in the notch. The intermodulation power content inside this notch or test channel follows a stochastic process, as the excitation itself is not deterministic. The NPR is a metric, which describes the distortions introduced by the TWT and is defined as [60]

$$\mathrm{NPR} = \frac{\mathrm{E}\left\{\int\limits_{-B_{\mathrm{Notch}}/2}^{B_{\mathrm{Notch}}/2} S_2(\omega-\omega_0)\,\mathrm{d}\omega\right\}}{\mathrm{E}\left\{\int\limits_{-B_{\mathrm{Notch}}/2}^{B_{\mathrm{Notch}}/2} S_2(\omega-\omega_{\mathrm{ref}})\,\mathrm{d}\omega\right\}} = \frac{\mathbf{E}\left\{P_{2,\omega_0}\right\}}{\mathbf{E}\left\{P_{2,\omega_{\mathrm{ref}}}\right\}}, \tag{4.2}$$

where $\mathbf{E}\{X\}$ is the expected value of a random variable X, P_{2,ω_0} denotes the portion of the output power inside the notch and $P_{2,\omega_{\text{ref}}}$ the power in a reference band outside the notch. This parameter represents a limit both for the distortions of complex modulations and for the intermodulation of multi-tone signals. The NPR can be calculated in SIMBA using both frequency-independent and frequency-dependent Envelope models. The calculation duration for the latter can sometimes be rather long. As described above, this can be helpful in the analysis, but offers hardly any hints for a more evolved synthesis.

In order to remedy this situation, approximate expressions are developed in the following, using approaches that are comparable to Envelope methods. These should not only enable a fast calculation of the NPR, but also provide insight into the dependencies of classical quantities. Since this happens at a model level, it is now primarily a mathematical task. Even for a few tones in the excitation, the nonlinearity of the TWT produces an infinite number of intermodulation products. A basically continuous spectrum, which has to be considered as excitation for the NPR, is at first much more complex. Nevertheless, the problem can be simplified by taking various assumptions into account. The full derivation consists of many purely algebraic steps, which do not add value to understanding. In the following, these steps are cut to present the derivation in a condensed form.

The calculation is carried out in two parts. For the numerator, we need to distinguish the individual frequency components at the output. Therefor, an approach using $2N$ CW carriers given by

$$x_1(t) = \frac{A_0}{\sqrt{2N}} \cdot \sum_{n \in \mathcal{U}} \cos(\omega_0 t + n\Delta\omega t + \varphi_n), \quad \mathcal{U} = [-N, N] \setminus \{0\}, \tag{4.3}$$

$$= \frac{2A_0}{\sqrt{2N}} \cdot \sum_{n=1}^{N} \cos\Bigg(\omega_0 t + \underbrace{\frac{\varphi_n + \varphi_{-n}}{2}}_{=:\overline{\varphi}_n}\Bigg) \cdot \cos\Bigg(n\Delta\omega t + \underbrace{\frac{\varphi_n - \varphi_{-n}}{2}}_{=:\underline{\varphi}_n}\Bigg), \tag{4.4}$$

with amplitude A_0, frequency $\omega_0 + n\Delta\omega$ and random phase distributions $\varphi_i \in \text{unif}(0, 2\pi)$, is chosen for the input signal, where the center carrier is not excited. This assumption has, for some purposes, shown to be a useful representation of the band-limited, notch-filtered input signal for which the NPR is defined. With $\Delta\omega \ll \omega_0$, the cosine terms at $n\Delta\omega$ change slowly compared to the one at ω_0, and can each be considered as an amplitude $A(t, n)$ which slowly changes over time. Thus, further sorting yields[2]

$$x_1(t) = \frac{2A_0}{\sqrt{2N}} \cdot \sum_{n=1}^{N} A(t,n)\cos(\omega_0 t + \overline{\varphi}_n) = \frac{2A_0}{\sqrt{2N}} |z| \cdot \cos(\omega_0 t + \arg(z)), \tag{4.5}$$

where $z = |z| \cdot \exp(j \arg(z))$ can be expressed by

$$|z| = \sqrt{\left(\sum_{n=1}^{N} \cos\left(n\Delta\omega t + \underline{\varphi}_n\right)\cos(\overline{\varphi}_n)\right)^2 + \left(\sum_{n=1}^{N} \cos\left(n\Delta\omega t + \underline{\varphi}_n\right)\sin(\overline{\varphi}_n)\right)^2} \tag{4.6}$$

and

$$\arg(z) = \arctan\left(\frac{\sum_{n=1}^{N} \cos\left(n\Delta\omega t + \underline{\varphi}_n\right)\sin(\overline{\varphi}_n)}{\sum_{n=1}^{N} \cos\left(n\Delta\omega t + \underline{\varphi}_n\right)\cos(\overline{\varphi}_n)}\right) =: \varphi_z(t). \tag{4.7}$$

This helps forming the analytic description considering the slowness of the cosine terms at $n\Delta\omega$

$$x_1(t) = \hat{x}_1(t)\cos(\omega_0 t + \varphi_z(t)), \tag{4.8}$$

such that, with Equation (4.4) and Equation (4.8), we can define the transforms

$$\left.\begin{matrix}\cos(\omega_0 t + \varphi_z(t)) \\ \sin(\omega_0 t + \varphi_z(t))\end{matrix}\right\} \cdot \hat{x}_1(t) = \frac{2A_0}{\sqrt{2N}} \sum_{n=1}^{N} \cos(n\Delta\omega t + \underline{\varphi}_n) \cdot \left\{\begin{matrix}\cos\left(\omega_0 t + \overline{\varphi}_n\right) \\ \sin\left(\omega_0 t + \overline{\varphi}_n\right).\end{matrix}\right. \tag{4.9}$$

Using an expansion around the operating point [38][3] and analogously to other applications in [37], [38], one can approximate the signal x_2 at the output of the TWT, in case the signal amplitude is small compared to the saturated amplitude, by

$$x_2(t) \approx \hat{x}_1(t) G \left(1 - \frac{\lambda_s \hat{x}_1^2(t)}{2\hat{x}_{1,\mathrm{sat}}^2}\right) \cdot \cos\left(\omega_0 t + \varphi_z(t) + \lambda_\varphi \frac{\hat{x}_1^2(t)}{\hat{x}_{1,\mathrm{sat}}^2}\right), \tag{4.10}$$

where G denotes the stationary gain at the operating point, and the time-varying deviations of the gain and the nonlinear phase shift are considered by a drive-dependent power proportionality factor λ_s and a drive-dependent phase proportionality factor λ_φ.

Considering $\hat{x}_1(t) \ll \hat{x}_{1,\mathrm{sat}}$ for a small-angle approximation and with Equation (4.9), we can rewrite Equation (4.10) as

$$x_2(t) \approx \hat{x}_1(t) G \left(1 - \frac{\lambda_s \hat{x}_1^2(t)}{2\hat{x}_{1,\mathrm{sat}}^2}\right) \left[\cos(\omega_0 t + \varphi_z(t)) - \frac{\lambda_\varphi \hat{x}_1^2(t)}{\hat{x}_{1,\mathrm{sat}}^2} \sin(\omega_0 t + \varphi_z(t))\right] \tag{4.11}$$

$$= \frac{2A_0 G}{\sqrt{2N}} \cdot \left(A_1(t) - \lambda_\varphi A_2(t) - \lambda_s A_3(t) + \lambda_\varphi \lambda_s A_4(t)\right). \tag{4.12}$$

[2] Generally speaking, we can always express the superposition of multiple phase shifted sinusoidal signals at frequency ω_0 as another sinusoidal signal at frequency ω_0.

[3] According to [38], the origin of the expansion is adapted from [68], which unfortunately is not available.

with

$$A_1(t) = \sum_{n=1}^{N} \cos\left(\omega_0 t + \overline{\varphi}_n\right) \cos(n\Delta\omega t + \underline{\varphi}_n) \tag{4.13}$$

describing an exclusively scalar amplification of the initial frequency components of $x_1(t)$, not having any components at ω_0, and

$$A_2(t) = \sum_{n=1}^{N} \frac{\hat{x}_1^2(t)}{\hat{x}_{1,\text{sat}}^2} \sin\left(\omega_0 t + \overline{\varphi}_n\right) \cos(n\Delta\omega t + \underline{\varphi}_n), \tag{4.14}$$

$$A_3(t) = \sum_{n=1}^{N} \frac{\hat{x}_1^2(t)}{2\hat{x}_{1,\text{sat}}^2} \cos\left(\omega_0 t + \overline{\varphi}_n\right) \cos(n\Delta\omega t + \underline{\varphi}_n), \text{ and} \tag{4.15}$$

$$A_4(t) = \sum_{n=1}^{N} \frac{\hat{x}_1^4(t)}{2\hat{x}_{1,\text{sat}}^4} \sin\left(\omega_0 t + \overline{\varphi}_n\right) \cos(n\Delta\omega t + \underline{\varphi}_n), \tag{4.16}$$

including frequency components due to intermodulation. They arise from the multiplication of $\cos(n\Delta\omega t + \underline{\varphi}_n)$ with $\hat{x}_1^2(t)$. As afterwards no further mixing is done, it is from now on sufficient to consider the A_{i,ω_0} instead of the A_i. The A_{i,ω_0} denote all components of the A_i at ω_0. This leads to

$$\begin{aligned} A_{2,\omega_0}(t) = \frac{2A_0^2}{4N\hat{x}_{1,\text{sat}}^2} \sum_{\mathcal{X}} \Big[&\sin\left(\omega_0 t + \overline{\varphi}_n\right) \\ &\cdot \cos(\overline{\varphi}_{|m|} - \overline{\varphi}_{|p|}) \cos\left(\underline{\varphi}_n + \epsilon(m) \cdot \underline{\varphi}_{|m|} + \epsilon(p) \cdot \underline{\varphi}_{|p|}\right) \Big], \end{aligned} \tag{4.17}$$

where

$$\mathcal{X} := \left\{ \begin{matrix} n \in [1, N], \\ m, p \in [-N, N] \backslash \{0\} \end{matrix} \;\middle|\; n + m + p = 0 \right\}, \tag{4.18}$$

and

$$\epsilon(x) := \frac{x}{|x|}, \quad \text{for } x \neq 0. \tag{4.19}$$

An analytic signal description according to Equation (4.5)-Equation (4.7) yields

$$A_{2,\omega_0}(t) = \frac{A_0^2}{2N\hat{x}_{1,\text{sat}}^2} |z_2| \cdot \sin(\omega_0 t + \arg(z_2)), \tag{4.20}$$

where

$$|z_i| = \sqrt{\left(A_{z,i}^{(I)}\right)^2 + \left(A_{z,i}^{(Q)}\right)^2}, \quad \varphi_{z,i} := \arg(z_i) = \arctan\left(\frac{A_{z,i}^{(Q)}}{A_{z,i}^{(I)}}\right), \tag{4.21}$$

with

$$\left.\begin{matrix} A_{z,2}^{(I)} \\ A_{z,2}^{(Q)} \end{matrix}\right\} := \sum_{\mathcal{X}} \cos(\overline{\varphi}_{|m|} - \overline{\varphi}_{|p|}) \cos\left(\underline{\varphi}_n + \epsilon(m) \cdot \underline{\varphi}_{|m|} + \epsilon(p) \cdot \underline{\varphi}_{|p|}\right) \left\{ \begin{matrix} \cos(\overline{\varphi}_n) \\ \sin(\overline{\varphi}_n). \end{matrix} \right. \tag{4.22}$$

Analogously for the other amplitude factors, one can find

$$A_{3,\omega_0}(t) = \frac{A_0^2}{4N\hat{x}_{1,\text{sat}}^2}\,|z_3| \cdot \cos(\omega_0 t + \arg(z_3)), \tag{4.23}$$

with $z_3 = z_2$, and finally, for the last term,

$$A_{4,\omega_0}(t) = \frac{A_0^4}{8N^2\hat{x}_{1,\text{sat}}^4}\,|z_4| \cdot \sin(\omega_0 t + \arg(z_4)) \tag{4.24}$$

with

$$\left.\begin{matrix} A_{z,4}^{(I)} \\ A_{z,4}^{(Q)} \end{matrix}\right\} := \sum_{\mathcal{Y}} B(n,m,p,q,r) \cdot \begin{cases} \cos(\overline{\varphi}_n), \\ \sin(\overline{\varphi}_n), \end{cases} \tag{4.25}$$

where $B(n,m,p,q,r)$ is a function of phase relations similar to (but more involved than) the one in Equation (4.17), and

$$\mathcal{Y} := \left\{ \begin{matrix} n \in [1,N], \\ m,p,q,r \in [-N,N]\backslash\{0\} \end{matrix} \;\middle|\; n+m+p+q+r=0 \right\}. \tag{4.26}$$

$A_{z,2}^{(I,Q)}$ and $A_{z,4}^{(I,Q)}$ only depend on the number of considered carriers $2N$, A_0 denotes the operating point, and $\lambda_\varphi, \lambda_s, G$ and $\hat{x}_{1,\text{sat}}(t)$ are partly operating-point dependent characteristics of the TWT. The proportionality factors λ_s (power) and λ_φ (phase) can be approximated by the known TWT transfer-curve characteristics, i.e., the compression c and k_p [38]

$$c \approx \lambda_s \frac{A_0^2}{2\hat{x}_{1,\text{sat}}^2}, \tag{4.27}$$

and

$$k_\text{p} \approx \lambda_\varphi \frac{A_0^2}{2\hat{x}_{1,\text{sat}}^2}, \tag{4.28}$$

such that Equation (4.17) can now be represented by TWT-independent constants and the two known TWT characteristics. Thus, the power in the notch results in

$$P_{2,\omega_0} = \frac{A_0^2 G^2}{2} \cdot \left(a_1 \frac{c^2 k_\text{p}^2}{N^5} + a_2 \frac{c^2 k_\text{p}}{N^4} + a_3 \frac{c^2}{N^3} + a_4 \frac{c k_\text{p}^2}{N^4} + a_5 \frac{c k_\text{p}}{N^3} + a_6 \frac{k_\text{p}^2}{N^3} \right). \tag{4.29}$$

For the numerator of Equation (4.2), we need to calculate the expected value of Equation (4.29). The six parameters $\mathbf{E}\{a_j\}$ $(j \in [1,6])$ all consist of $A_{z,2}^{(I,Q)}$ and $A_{z,4}^{(I,Q)}$, which in turn consist of combinations of the initial $2N$ uniformly distributed random variables φ_i $(i \in \mathcal{U})$. For large N, the $A_{z,2}^{(I,Q)}$ and $A_{z,4}^{(I,Q)}$ will follow zero-mean normal distributions. They are not fully independent from each other, but only weakly correlated, with $\mathrm{Cov}\{A_{z,2}^{(I,Q)}, A_{z,4}^{(I,Q)}\} \approx 0$. Therefore, one can simplify

$$\mathbf{E}\{a_1\} = \mathbf{E}\left\{\frac{1}{4}\left(\left(A_{z,4}^{(I)}\right)^2 + \left(A_{z,4}^{(Q)}\right)^2\right)\right\} = \frac{1}{4}\mathrm{Var}\left\{A_{z,4}^{(I)}\right\} + \frac{1}{4}\mathrm{Var}\left\{A_{z,4}^{(Q)}\right\}, \tag{4.30}$$

$$\mathbf{E}\{a_2\} = \mathbf{E}\left\{\frac{1}{2}\left(A_{z,2}^{(I)}A_{z,4}^{(Q)} - A_{z,2}^{(Q)}A_{z,4}^{(I)}\right)\right\} \approx 0, \tag{4.31}$$

$$\mathbf{E}\{a_3\} = \mathbf{E}\left\{\frac{1}{4}\left(\left(A_{z,2}^{(I)}\right)^2 + \left(A_{z,2}^{(Q)}\right)^2\right)\right\} = \frac{1}{4}\mathrm{Var}\left\{A_{z,2}^{(Q)}\right\} + \frac{1}{4}\mathrm{Var}\left\{A_{z,2}^{(I)}\right\}, \tag{4.32}$$

$$\mathbf{E}\{a_4\} = \mathbf{E}\left\{A_{z,2}^{(Q)}A_{z,4}^{(Q)} - A_{z,2}^{(I)}A_{z,4}^{(I)}\right\} \approx 0, \tag{4.33}$$

$$\mathbf{E}\{a_5\} = 0, \quad \text{and} \tag{4.34}$$

$$\mathbf{E}\{a_6\} = 4\mathbf{E}\{a_3\} = \mathrm{Var}\left\{A_{z,2}^{(Q)}\right\} + \mathrm{Var}\left\{A_{z,2}^{(I)}\right\}. \tag{4.35}$$

For the denominator of Equation (4.2), the signal $x_1(t)$ is directly considered as a stochastic signal

$$x_1(t) = R \cdot \cos\left(\omega_0 t + \Phi_R\right), \tag{4.36}$$

where R is, e.g., a Rayleigh distributed process, while Φ_R follows a uniform distribution in $\mathrm{unif}(0, 2\pi)$[4]. Analogously to Equation (4.10), we can now estimate the output signal

$$x_2(t) \approx RG\left(1 - \lambda_s \frac{R^2}{2\hat{x}_{1,\mathrm{sat}}^2}\right) \cdot \cos\left(\omega_0 t + \Phi_R + \lambda_\varphi \frac{R^2}{\hat{x}_{1,\mathrm{sat}}^2}\right) \tag{4.37}$$

$$= RG\left(1 - \lambda_s \frac{R^2}{2\hat{x}_{1,\mathrm{sat}}^2}\right)\sqrt{1 + \lambda_\varphi^2 \frac{R^4}{\hat{x}_{1,\mathrm{sat}}^4}} \cos\left(\omega_0 t + \arctan\left(\frac{\sin(\Phi_R) + \lambda_\varphi \frac{R^2}{\hat{x}_{1,\mathrm{sat}}^2}\cos(\Phi_R)}{\cos(\Phi_R) - \lambda_\varphi \frac{R^2}{\hat{x}_{1,\mathrm{sat}}^2}\sin(\Phi_R)}\right)\right). \tag{4.38}$$

Then, finally, the expected value of the total power at the output of the TWT is

$$\mathbf{E}\{P_2\} = \mathbf{E}\left\{|x_2|^2\right\} \tag{4.39}$$

$$= \mathbf{E}\left\{\frac{R^2G^2}{2}\left(1 - 2\frac{\lambda_s R^2}{2\hat{x}_{1,\mathrm{sat}}^2} + \frac{\lambda_s^2 R^4}{4\hat{x}_{1,\mathrm{sat}}^4}\right)\left(1 + \lambda_\varphi^2 \frac{R^4}{\hat{x}_{1,\mathrm{sat}}^4}\right)\right\}. \tag{4.40}$$

Here, we can use the approximate relationships

$$c \approx \lambda_s \frac{R^2}{\hat{x}_{1,\mathrm{sat}}^2}, \quad k_\mathrm{p} \approx \lambda_\varphi \frac{R^2}{\hat{x}_{1,\mathrm{sat}}^2}, \tag{4.41}$$

which then yield

$$\mathbf{E}\{P_2\} \approx \sigma^2 G^2 \left(1 - \frac{c}{2}\right)^2 \left(1 + k_\mathrm{p}^2\right), \tag{4.42}$$

[4]The calculation does not require knowledge of the full distribution, but only of some of its moments.

where σ is the scale parameter of the distribution R. In accordance with Equation (4.3), it is chosen to $\sigma = A_0/\sqrt{2}$. Thus, the NPR can be estimated from

$$\mathrm{NPR} = \frac{\mathbf{E}\left\{P_{2,\omega_0}\right\}}{\mathbf{E}\left\{P_{2,\omega_{\mathrm{ref}}}\right\}} = \frac{2N \cdot \mathbf{E}\left\{P_{2,\omega_0}\right\}}{\mathbf{E}\left\{P_2\right\}} \approx \frac{b_1 c^2 k_{\mathrm{p}}^2 + b_2\left(c^2 + 4k_{\mathrm{p}}^2\right)}{\left(1 - \frac{c}{2}\right)^2\left(1 + k_{\mathrm{p}}^2\right)}, \tag{4.43}$$

which only depends on the two (fixed) scalars

$$b_1 := \frac{\mathrm{Var}\left\{A_{z,4}^{(I)}\right\} + \mathrm{Var}\left\{A_{z,4}^{(Q)}\right\}}{4N^4} \approx 50.97 \tag{4.44}$$

and

$$b_2 := \frac{2\mathrm{Var}\left\{A_{z,2}^{(I)}\right\} + 2\mathrm{Var}\left\{A_{z,2}^{(Q)}\right\}}{N^2} \approx 0.358 \tag{4.45}$$

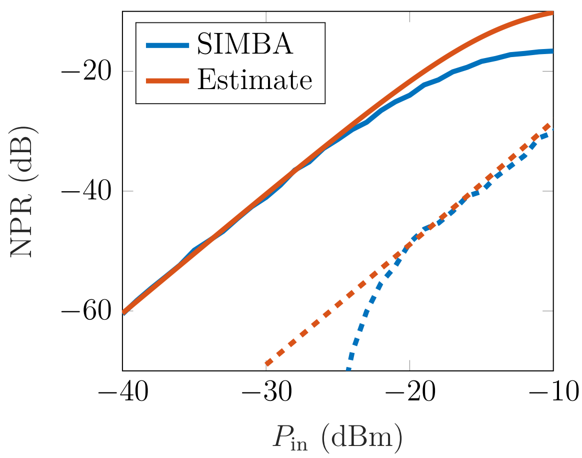
and the two drive-dependent amplifier parameters c and k_{p}.

Again, considering a reference Ku-Band TWT, a comparison between the approximate solution and simulation in SIMBA as shown in Figure 4.8a reports good agreement far in back-off. This is especially the case when no phase distortion is considered ($k_p = 0$), but also with $k_p \neq 0$ the expression yields close agreement up to relevant back-off regions of around 10 dB in input back-off. The variation far in back-off in the case without phase distortion stems from a simplification in the simulation with SIMBA. For this, the nonlinear gain compression is related to the gain as observed with a small-signal reference simulation. Thus, in SIMBA it starts at zero for a small-signal input power level, while the discussion in Section 2.3 hints at the occurrence of a nonlinear gain already in a purely ballistic setting. Theoretically, the gain compression drops continuously towards zero-drive excitation, causing intermodulation at any non-zero drive level.

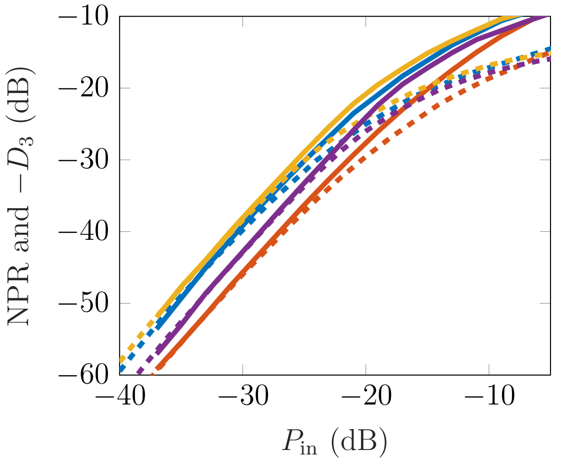
Comparing the expression in Equation (4.43) to the third order intermodulation ratio approximation from Equation (3.20), a relationship

$$\mathrm{NPR} \approx \frac{b_1 c^2 k_{\mathrm{p}}^2 + b_2\left(c^2 + 4k_{\mathrm{p}}^2\right)}{\left(1 - \frac{c}{2}\right)^2\left(1 + k_{\mathrm{p}}^2\right)} \propto 10^{0.1 D_3} \approx \left(c^2 + 4k_{\mathrm{p}}^2\right) \tag{4.46}$$

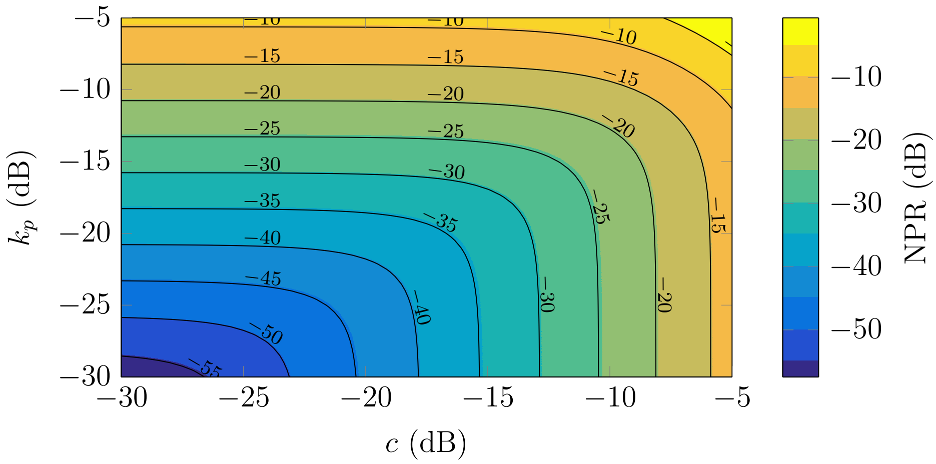
is found for small c and k_{p}. Thus, the D_3 can be translated to the NPR in logarithmic scale by a constant, TWT-independent shift, when considering small input powers. In Figure 4.8b, this is shown for four different AM-AM and AM-PM curve sets, serving as a reference for the Envelope simulation of D_3 and NPR in SIMBA. For this purpose, an MVTRAD CW simulation is carried out for a large range of input powers, starting from input powers of around 50 dB below saturation. Also, a fine sampling along



(a) Comparison with (solid line) and without (dashed line) phase distortion.



(b) NPR with constant shift (dashed line) compared to D_3 (solid line) for four different transfer-curve sets.



(c) Contour plot of the NPR as a function of c and k_p.

Figure 4.8: NPR results from estimate expression.

the input power range is chosen, such that the reference curves are accurate and smooth and can be interpolated with close to no error. The corresponding results from the same reference data are denoted by equal coloring. Far in back-off the slopes correspond well. As expected, this agreement worsens, the closer to saturation the operating point is chosen.

At last, Figure 4.8c shows the impact of k_p and c on the expected NPR. It can be seen that a large improvement on, e.g., c does not help much, when k_p is already large, and vice-versa.

4.2.2 Frequency-Dependent Intermodulation Ratio

In Section 3.3.1, a simple two-tone expression from literature has been reviewed, approximating the third-order intermodulation ratio for two equally strong tones. It was conceived assuming constant transfer behavior across the considered input tones and intermodulation products. Such a simplified expression can serve as a quick reference and help understanding the effect of the individual nonlinear effects. To extend this to a more broadband expression, the derivation is incorporated into a TBM-like approach as introduced in Section 3.3.4. For this purpose, a two-tone signal

$$x(t) = A_1 \sin\left(\left(\omega_0 - \frac{\omega_\Delta}{2}\right) t\right) + A_2 \sin\left(\left(\omega_0 + \frac{\omega_\Delta}{2}\right) t\right) \tag{4.47}$$

with similar amplitudes $A_1 \approx A_2$ and close frequency spacing $\omega_\Delta \ll \omega_0$ is considered. Constant phase terms, unnecessary for the resulting expression, are neglected throughout this derivation. The TBM approach consists of two linear filters framing the nonlinearities. At the end of the first filter, a signal

$$y_1(t) = A_1 G_\Delta\left(-\frac{\omega_\Delta}{2}\right) \sin\left(\left(\omega_0 - \frac{\omega_\Delta}{2}\right) t\right) + A_2 G_\Delta\left(\frac{\omega_\Delta}{2}\right) \sin\left(\left(\omega_0 + \frac{\omega_\Delta}{2}\right) t\right) \tag{4.48}$$

$$= A_1 A_2 A_\Delta(t) \sin\left(\left(\omega_0 - \frac{\omega_\Delta}{2}\right) t + \varphi_\Delta\right) \tag{4.49}$$

is obtained, with

$$A_\Delta(t) = \sqrt{\frac{G_\Delta^2\left(-\frac{\omega_\Delta}{2}\right)}{A_2^2} + \frac{G_\Delta^2\left(\frac{\omega_\Delta}{2}\right)}{A_1^2} + \frac{2G_\Delta\left(\frac{\omega_\Delta}{2}\right) G_\Delta\left(-\frac{\omega_\Delta}{2}\right)}{A_1 A_2} \cos(\omega_\Delta t)} \tag{4.50}$$

and

$$\varphi_\Delta(t) = \arctan\left(\frac{A_2 G_\Delta\left(\frac{\omega_\Delta}{2}\right) \sin(\omega_\Delta t)}{A_2 G_\Delta\left(\frac{\omega_\Delta}{2}\right) \cos(\omega_\Delta t) + A_1 G_\Delta\left(-\frac{\omega_\Delta}{2}\right)}\right), \tag{4.51}$$

where $G_\Delta(\omega)$ is the impact of the input filter compensating for the different degree of nonlinearity seen at a frequency ω compared to the assumed reference center frequency

ω_0. Analogously to Equation (4.10), the nonlinear transfer of the reference curve at ω_0 is applied to $y_1(t)$, resulting in a signal $y_2(t)$ at the output of the nonlinear block. Considering trigonometric identities, a small-angle approximation, and neglecting small terms, the components of interest are be extracted, where

$$A_{2,\left(\omega_0-\frac{3}{2}\omega_\Delta\right)} \approx GA_2G_\Delta\left(\frac{\omega_\Delta}{2}\right)G_\Delta^2\left(-\frac{\omega_\Delta}{2}\right)\frac{A_1^2}{A_{\text{sat}}^2}\sqrt{\frac{\lambda_s^2}{4}+\lambda_\varphi^2} \tag{4.52}$$

$$\approx GG_\Delta\left(\frac{\omega_\Delta}{2}\right)G_\Delta^2\left(-\frac{\omega_\Delta}{2}\right)A_2\sqrt{\left(\frac{c}{2}\right)^2+k_p^2} \tag{4.53}$$

is the amplitude at $\omega_0-\frac{3}{2}\omega_\Delta$ and

$$A_{2,\left(\omega_0+\frac{3}{2}\omega_\Delta\right)} \approx GA_1G_\Delta\left(-\frac{\omega_\Delta}{2}\right)G_\Delta^2\left(\frac{\omega_\Delta}{2}\right)\frac{A_2^2}{A_{\text{sat}}^2}\sqrt{\frac{\lambda_s^2}{4}+\lambda_\varphi^2} \tag{4.54}$$

$$\approx GG_\Delta\left(-\frac{\omega_\Delta}{2}\right)G_\Delta^2\left(\frac{\omega_\Delta}{2}\right)A_1\sqrt{\left(\frac{c}{2}\right)^2+k_p^2} \tag{4.55}$$

is the amplitude at $\omega_0+\frac{3}{2}\omega_\Delta$. Similarly to Section 4.2.1, G denotes the stationary gain and the time-varying deviations of the gain and nonlinear phase shift are considered by the proportionality factors λ_s (power) and λ_φ (phase). For $D_3^{(1,2)}$, also the powers at the input frequencies are required. For simplicity, it can be assumed that at these frequencies the linear amplification outweighs higher order intermodulation products. Assuming $A_1{=}A_2$, the third order intermodulation ratio is calculated from

$$D_3^{(1)} \approx D_{3,0} - 20\log_{10}\left(\frac{G_\Delta\left(\frac{\omega_\Delta}{2}\right)G_\Delta\left(-\frac{\omega_\Delta}{2}\right)H_\Delta\left(-\frac{3\omega_\Delta}{2}\right)}{H_\Delta\left(-\frac{\omega_\Delta}{2}\right)}\right), \tag{4.56}$$

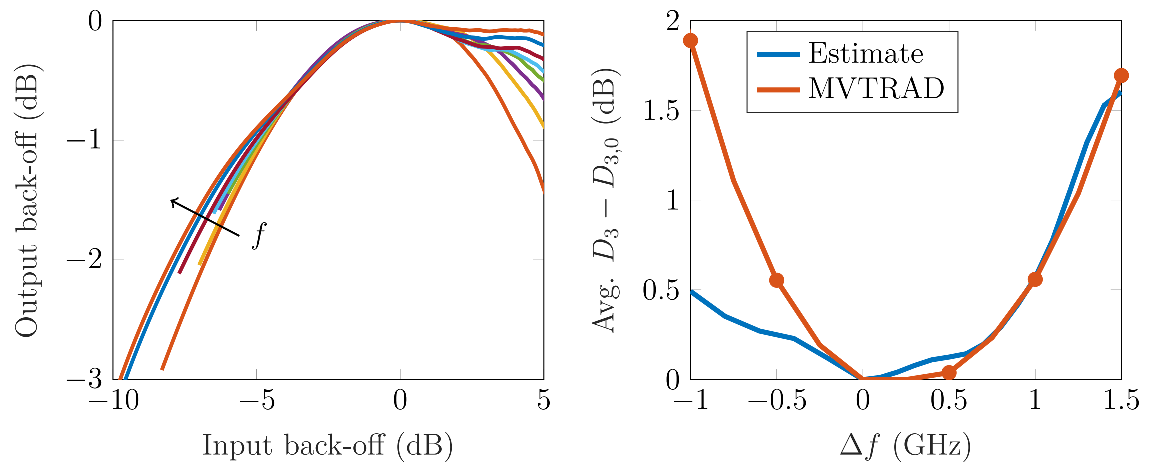
and

$$D_3^{(2)} \approx D_{3,0} - 20\log_{10}\left(\frac{G_\Delta\left(\frac{\omega_\Delta}{2}\right)G_\Delta\left(-\frac{\omega_\Delta}{2}\right)H_\Delta\left(\frac{3\omega_\Delta}{2}\right)}{H_\Delta\left(\frac{\omega_\Delta}{2}\right)}\right), \tag{4.57}$$

with

$$D_{3,0} = -20\log_{10}\left(\sqrt{\left(\frac{c}{2}\right)^2+k_p^2}\right), \tag{4.58}$$

following the definition from Equation (3.8). $H_\Delta(\omega)$ is the equivalent filter at the output, compensating for the difference in output power seen at ω compared to ω_0 for the same drive level. It can be seen that for both $D_3^{(1,2)}$ an additive term is to be considered compared to the static expression in Equation (3.20). This additional term compensates for the difference in back-off and gain. Also, as expected, the two ratios are different, resulting in an asymmetric distribution.



(a) Normalized transfer curves over frequency.

(b) Average $D_3 - D_{3,0}$ from the approximate expressions.

Figure 4.9: TWT data from MVTRAD and intermodulation product results.

Figure 4.9a shows the AM-AM curves generated in MVTRAD for a reference Ku-Band TWT. The curves are shifted by their saturation point, such that the shape similarity over frequency in the band of interest can be seen. The expressions from Equations (4.56) to (4.58) are evaluated for this reference TWT in comparison to two-tone MVTRAD simulation. Figure 4.9b shows the results, when one tone is fixed at the band center, while the other is placed at a distance Δf from it. As the expression is not valid for drive levels close to saturation, a realistic operating point 3 dB in output back-off is chosen. It can be seen that the frequency-dependence of the broadband dual-tone intermodulation is well matched in the approximate expression even up to a large frequency difference of 1.5 GHz in positive direction. For negative Δf, this does not hold similarly well, which can directly be understood from the shape variations of the transfer curves, which are more pronounced in this frequency range. Thus, to apply the derived expressions, the transfer curves themselves need to be thoroughly evaluated.

This is also one reason why the derivation of similar expressions for the estimation of k_T is not feasible with reasonable effort. The variations of phase-related properties over frequency are significantly larger, not only considering an absolute shift, but also in the behavior over drive. This only allows for frequency-dependent solutions, where the considered band is rather narrow, rendering the resulting expressions more or less useless. Similarly, this also holds for broadband NPR estimations, which are considerably more demanding in terms of bandwidth and linearity.

4.3 Automated Back-Off Optimization

Although simple expressions to connect CW characteristics with ones from modulated operation have been found, the interpretation of these expressions with respect to the previously mentioned multidimensional problem with its many trade-offs still presents an ambitious endeavor. An alternative, heuristic approach is found analogously to the practical design of a delay line in the CW optimization of the TWT. Here, a satisfactory solution for the structure is iteratively calculated according to a previously defined target function.

The amplifier performance strongly relies on the pitch profile of the delay line, both for CW and modulated operation. As one cannot find a simultaneous optimum in all characteristics and no analytic large-signal formulas are available, no obvious solution exists. In this context, manual optimization is fairly time- and therefore resource-consuming. Thus, in the past, various automated optimization routines have been developed [69]–[71]. For this purpose, algorithms based on Simulated Annealing (SA, [72]) have shown to be especially valuable and therefore have been used in various taper optimization approaches. Classically, these approaches are based on single-tone requirements, as can be obtained by frequency-domain codes. These are, among others, the saturated input and output power, nonlinear phase, small-signal gain, compression, and frequency flatness. Lately, optimization tools taking the device nonlinearity and thus the expected signal distortion into account have been published [73]. While fulfilling their aim of designs with improved linearity for modulated operation, they do not directly consider the effective efficiency, although it is the other main cause for the performance decline in back-off. There, especially the excess collector heat deteriorates the effective total efficiency.

In this work, which draws from previously published results in [74], CW goals, nonlinear distortions, and collector, as well as beam efficiencies in modulated operation are combined to enable SA taper optimization specifically for communication signals. The advantage is, of course, the possibility to achieve a context-dependent adaptation without much understanding of the relationships between the variables to be optimized and the adjustment screws of the structure. The underlying optimization routine chosen for this is based on a previously published one [75], where the basics of the SA approach and the CW optimization workflow, as well as the CW goal function are discussed in detail.

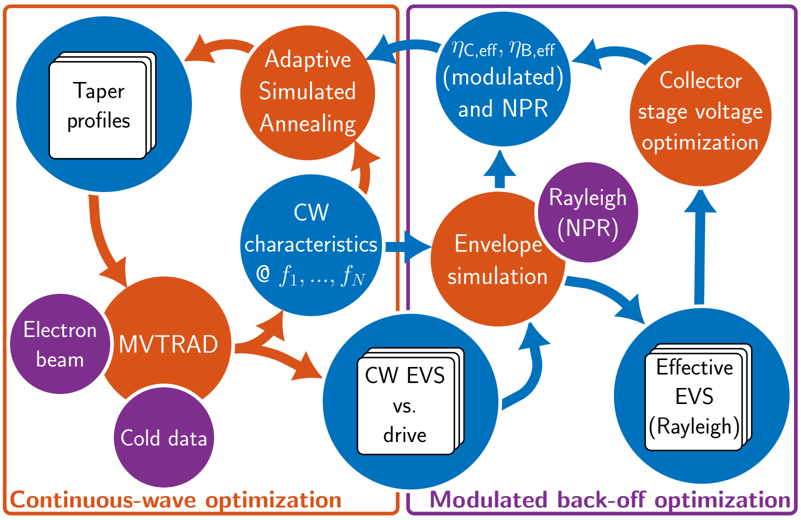


Figure 4.10: Workflow of the extended SA optimizer, which considers the modulated back-off.

4.3.1 Method and Workflow

In the following, the general approach is shortly outlined. In the classical helix-taper optimization, as is schematically shown in the left half of Figure 4.10, an initial taper is simulated, e.g., in MVTRAD, with regards to its CW characteristics. Therefor, a wide set of cold parameters is pregenerated, which covers the whole domain of possible tapers. Also, auxiliary information, such as the beam data, are predefined and invariant. To evaluate the performance against specified goals and adapt to the taper accordingly, an SA approach is used. The idea is that similarly to a cooling process of a melted material, the initially not-optimal state gradually "cools down" to a well-ordered, low-energy state, which in this case means an optimized taper. Therefor, in [75], an exponential cooling schedule

$$T_{n+1} = T_n \cdot \alpha = T_0 \cdot \alpha^n,\ \alpha \in [0, 1[\tag{4.59}$$

was chosen, which determines the temperature reduction from one iteration to another. According to the current temperature inside the SA code, the next taper then is generated by random, defined alterations of the current taper.

Here, the CW optimization is extended by an additional module, taking effective quantities into account. Therefor, the CW results are processed in an Envelope-based routine. Also, the EVS's are recorded for a range of drive levels. This way, effective values are calculated, which serve as an additional source for the goal function defined in the SA approach. This extended module is shown in Figure 4.10.

4.3.2 Target Functions

The optimization is realized by minimizing a total energy goal function

$$E(\text{current taper}) = \frac{1}{M}\sum_{j=1}^{M}\left[\frac{w_j}{N_j X_{\mathrm{T},j}^2}\sum_{i=1}^{N_j}\mathrm{H}(X_{\mathrm{G},i,j}-X_{\mathrm{T},j})\,(X_{\mathrm{G},i,j}-X_{\mathrm{T},j})^2\right], \quad (4.60)$$

which consists of a weighted error between the individual j-th partial goal of in total M goals, with weighting factors w_j, target values $X_{\mathrm{T},j}$, and the current performance value $X_{\mathrm{G},i,j}$. For each goal value, N_j frequency points are considered, which might differ according to the individual goal. $\mathrm{H}(\cdot)$ denotes the unit step function, ensuring that a goal which is fulfilled with a large margin does not obscure a different, missed goal.

This way, both CW and effective figures are integrated into a single goal function for the SA approach to decide on the performance of the taper. Individual weighting of the results enables highlighting certain aspects for the optimization. Details on the goal functions are given in the following.

Continuous-Wave Goals

The partial goal function for the CW goals consists of the typical amplifier characteristics for various frequencies, spread along the considered frequency band[5]. For each of these frequencies, the input and output power, the nonlinear phase shift at saturation, the linear gain, and the compression are recorded and evaluated against predefined target values. This is as defined in [75]. For frequency flatness and power symmetry about the band center, the results are compared at several frequencies.

Modulation-Based Goals

The extension of the optimizer towards modulated operation is enabled by appending an Envelope method to the MVTRAD simulation and including modulation-based goals to the SA algorithm. While in principle the broadband performance is of interest, and thus the QPM would be advantageous in this respect, it comes with a large cost in terms of computation time. In a multi-dimensional, complex, automated optimization, the time per iteration is highly critical. For quick evaluation of the taper, the SGC model is chosen. Its major limit here is given by the frequency flatness of the TWT characteristics. Luckily, the CW goal description ensures the validity of such a model

[5] In this case, three frequencies are chosen: two at the boundaries of the band and one in its center. While this is obvious for classical optimization, another choice could have advantages considering effective metrics.

for realistic bandwidths, enabling the usability of the SGC model for the proposed purpose.

In this approach, both effective beam and collector efficiency are considered. To obtain the effective collector efficiency in modulated operation, the EVS $I(V)$ is taken into account. Using MVTRAD, each taper is evaluated from small-signal drive to a few dB into overdrive to generate a set of CW $I_{\mathrm{CW}}(V, P_{\mathrm{in}})$ results. This data is then used to transform an envelope distribution into the corresponding effective EVS $I_{\mathrm{mod}}(V)$, which relates to a weighted average of the $I_{\mathrm{CW}}(V, P_{\mathrm{in}})$ and enables the estimation of effective collector efficiencies, as was outlined in Section 3.3.2.

The envelope distribution may vary significantly even in the scope of communication signals. Thus, the effective collector efficiency strongly depends on the chosen input signal modulation, which is not always known. To optimize the taper independently from the signal choice, a distribution representing a more general setting is desired. This is also the idea behind the NPR. The total signal from multiple superimposed modulated carriers is noise-like in both inphase and quadrature component, with the envelope distribution following a Rayleigh distribution. Its scale parameter, which is the only parameter in the distribution, is related to the average amplitude of the signal. For a fixed input or output back-off level, the procedure from Section 3.3.2 returns the effective $I_{\mathrm{mod}}(V)$ for such a Rayleigh distributed signal. Thus, hopefully, the optimizer also yields an effective electron spectrum, which shows better collectability compared to a CW approach.

In the end, the efficiency strongly depends on the choice of the collector stage voltages. For meaningful comparison between two designs, the respective best collector potentials need to be found for each of them. The first and last stage have a protective function for the device and are therefore left unchanged. Instead, for a four-stage design, the two remaining stages are optimized for the given effective $I_{\mathrm{mod}}(V)$. As the range of possible solutions, the dynamics of the function to be optimized, and the required potential accuracy are all rather low, even a simple brute-force approach suffices for this purpose. From the optimized stages, an effective collector efficiency $\eta_{\mathrm{Coll,eff}}$ is extracted.

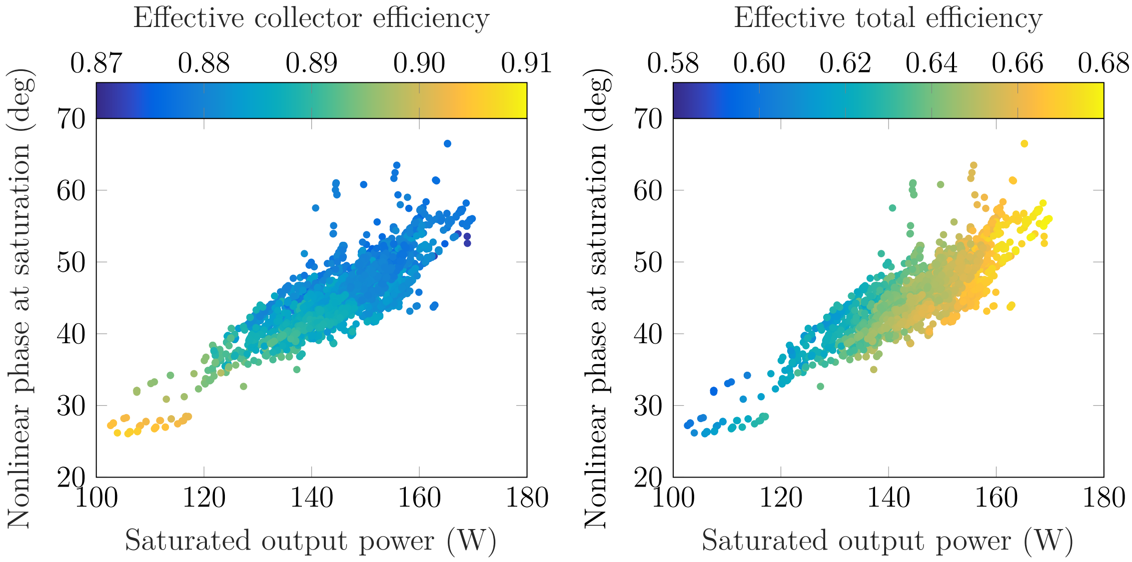
The same considerations regarding the envelope distribution also hold for the effective beam efficiency $\eta_{\mathrm{Beam,eff}}$. Again, for consistency, a Rayleigh distribution is considered to estimate the effective output power, from which the beam efficiency is calculated. Similarly, additional communication-related routines are implemented. For instance, the NPR is calculated using the Envelope approach from a Rayleigh distributed input signal. While the linearity can also be considered by means of the CW characteristics, the usage of the NPR enables a more direct understanding regarding the achieved degree of distortion reduction. As the Rayleigh distribution serves as the basis for the effective values, designs coming from this extended optimization routine are in the following referred to as a Rayleigh optimized designs.

4.3.3 Results

To evaluate the optimizer, a Ku-Band helix TWT from a pure CW optimization, in the following denoted as design (A), serves as the starting point for the proposed Rayleigh approach. Throughout an optimization run, a large set of data from a variety of tapers is typically generated and stored. This allows for some statistical investigation. In Figure 4.11, the effective collector and total efficiencies in back-off are shown for a selection of 3000 taper evaluations, each dot representing a single evaluation, and the color reporting the efficiency value at 3 dB in back-off. Some losses, such as coupler and filament losses are not considered, also the collector currents are estimated from the EVS using the described common approximation. For each taper, the saturated output power and phase are also recorded, such that the effective efficiencies are related to the CW goals at saturation. Small nonlinear phase shifts and high saturated output powers are desirable. Thus, the further down to the right a point is, the better are its CW properties. Unfortunately, the trade-off between these characteristics is easily visible, such that it is difficult to find such an optimal point. Instead, the evaluated points orient themselves roughly on a line from bottom left to the top right. Also the results considering the coloring of the dots are mostly as expected. A lower saturated output power or nonlinear phase shift typically means that the beam is not well spent. The remaining beam power is large and the collectability is high, leading to a high effective collector efficiency. As for the effective total efficiency, it includes the effective beam efficiency and therefore increases to the top right. Still, for a fixed output power, one can see an increased effective total efficiency in the direction of decreasing nonlinear phase shift, although the general trend is the opposite. This is a result of a better EVS, even though the same amount of power is extracted.

The Rayleigh optimization starting with design (A) leads to design (B), which is optimized for the effective collector efficiency $\eta_{\mathrm{Coll,eff}}$ and the effective total efficiency $\eta_{0,\mathrm{eff}}$ in back-off. Both $\eta_{0,\mathrm{eff}}$ and the CW total efficiency $\eta_{0,\mathrm{cw}}$ are estimated with $\eta_{\mathrm{Coll,eff}}$, $\eta_{\mathrm{Beam,eff}}$, $\eta_{\mathrm{Coll,cw}}$ and $\eta_{\mathrm{Beam,cw}}$, neglecting for instance coupler and heater losses, which are nearly identical for (A) and (B). Both designs fulfill all CW goals.

In Figure 4.12, the CW and effective efficiencies for both designs are shown. In Figure 4.12a, the collector efficiency is shown versus back-off. For CW excitation, it can be seen that although the saturated values are similar, but slightly higher for the CW

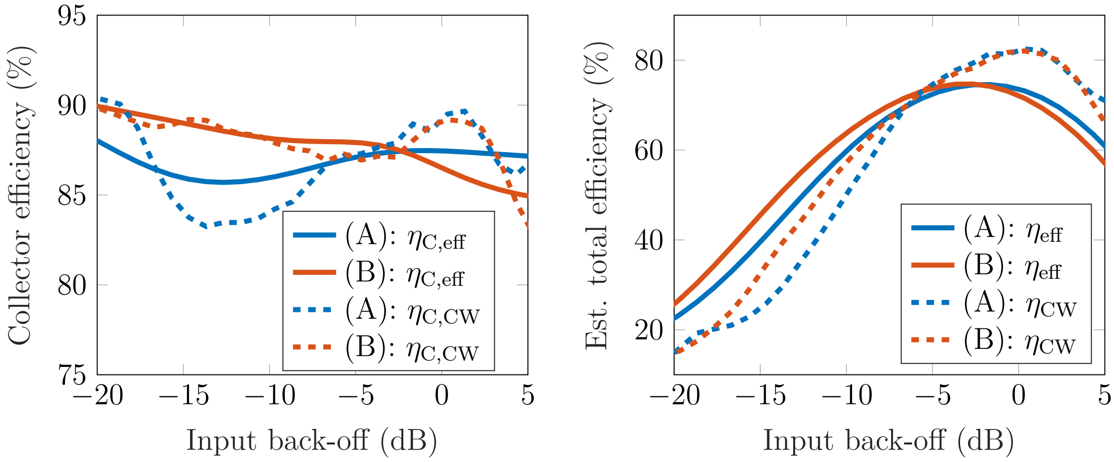


(a) Effective collector efficiency. **(b)** Effective total efficiency.

Figure 4.11: Effective back-off efficiencies for a selection of evaluated tapers versus CW output power and nonlinear phase shift.

optimized design (A), the collector efficiency in back-off strongly drops. This comes from the fact that the optimization is carried out, such that the effective efficiency is improved, leading to a more robust design than the singular point optimization in saturation. This can again be seen for the Rayleigh excited results, where the drop in back-off is in comparison flattened, but similarly well visible. This improvement directly impacts the total efficiency. As both designs fulfill the output power requirement, the beam efficiency is in consequence similar. The improved collector efficiency in back-off is clearly evident, increasing the total efficiency in simulation by a few percent. In total, for the effective values, a far more robust behavior over the drive is shown, with the efficiency-wise optimal operating point moved towards a linearity-wise preferred back-off point.

The exact improvement to be expected in reality cannot be determined directly and is most definitely lower than the shown differences, as it requires perfect knowledge of the spent beam over drive and how it behaves in the actual collector. A shortage in this knowledge leads to errors mainly in the collector efficiency, which is the main contributor to the improvement shown in Figure 4.12. For instance, the actual EVS is uncertain, e.g., due to inaccuracies in the magnetic focusing field and the lack of thermal electrons in the simulation, and can only be estimated. Also, a large contributor to uncertainty is the collector current calculation itself, as it relies on a common but



(a) Effective collector efficiency.

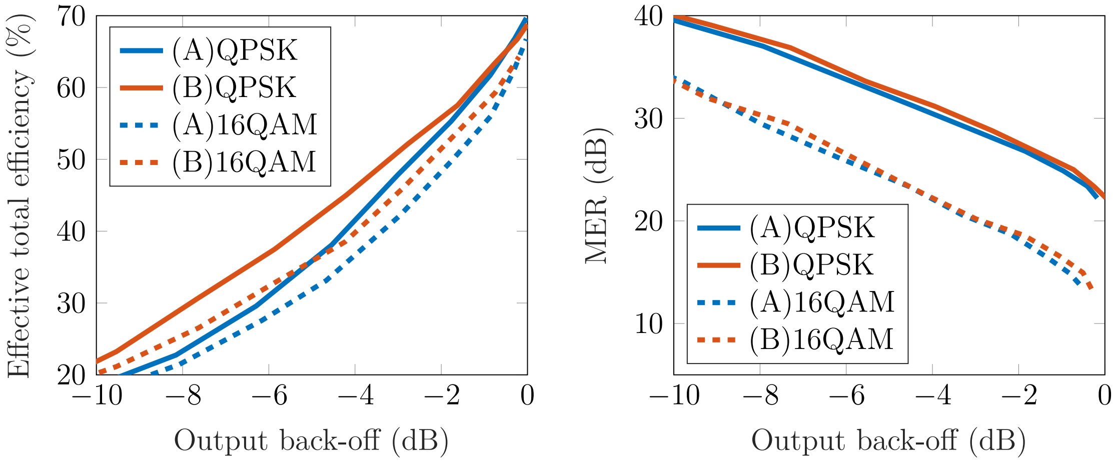
(b) Effective total efficiency.

Figure 4.12: Estimated CW and effective efficiencies of design (A) and (B) for an ideal four-stage collector, neglecting some loss mechanisms. CW excitation is shown by dashed lines, Rayleigh excitation by solid lines.

simplified procedure, assuming a perfectly sorting, symmetric multiple-stage collector without secondary electrons and other secondary effects [1]–[3]. While this has shown to be a valid approximation, it introduces some uncertainty in the optimum collector stage voltages, which contribute to all other results. In contrast, comparisons based on the nonlinearity and distortions are more trustworthy, as it was shown that the integral transfer properties are not strongly influenced by the expected typical beam uncertainties. Also, in contrast to the collector current estimation, no further sharp approximation is required.

To optimize against a distribution representing a generalized modulation setting, the Rayleigh distribution was chosen. To see the effect of the Rayleigh-based optimization on a known modulation scheme, Figure 4.13 shows the effective total efficiency and MER for a QPSK and a 16QAM signal, both versus output back-off. The improvement in back-off efficiency both in CW excitation and Rayleigh excitation also shows for the communication signals, the improvement being similarly large throughout the back-off. Luckily, the improved efficiency does not come at the cost of larger signal distortions, which can be seen in the MER, representing the device nonlinearity. In the relevant range around $-3\,\mathrm{dB}$ to $-5\,\mathrm{dB}$ output back-off, close to no change is seen for the 16QAM signal, and only around $0.5\,\mathrm{dB}$ difference for the QPSK signal.

In Section 4.1.2, some issues regarding the operating point choice and change according to the chosen modulation scheme were discussed. Here, a fixed output



(a) Effective total efficiency.

(b) Distortions.

Figure 4.13: Improvement of Rayleigh optimized design (B) compared to the CW optimized design (A), evaluated with communication signals.

back-off point is considered for the modulation goals. Together with the CW output power goal, this enables the comparison of the linearity throughout all evaluations. Predistortion is not considered here. Considering a linearizer would increase the complexity and uncertainty of the obtained results, as the real setup of linearizer and TWT differs from an ideal limiter, and the linearizer itself is typically generated based on TWT data. For instance, one might try to design the linearizer with the aim of minimizing phase distortions, possibly at the cost of the linear gain, and vice-versa. Depending on the operating point definition, collector stage number, linearization, and other auxiliary constraints, fundamentally different results and margins for improvement can be expected. Therefore, such a taper optimization is most promising when the actual application is well-defined, which unfortunately rarely is the case. Thus, the generalized Rayleigh approach, as it is considered here, helps obtaining structures, which cover improvements over a wide range of applications.

5 Conclusion

Traveling-wave tubes are high-efficiency vacuum electronic power amplifiers, which are predominantly used in satellite communications. Classical characterization is based on continuous-wave characteristics, such as the saturated input and output power, nonlinear phase, small-signal gain, compression, and frequency flatness. As a means to evaluate the linearity, additionally narrowband quantities such as the noise-power ratio, intermodulation product ratio, and phase transfer factor are evaluated.

Lately, modern satellite applications shift the modulation schemes from simple ones, e.g., constant-envelope quadrature phase-shift keying, to more complex amplitude- and phase-modulated signals. Thus, as more complex schemes typically are more sensitive to nonlinear distortions, the amplifier is typically operated further in the linear back-off region. This way, nonlinearity-induced modulation errors are diminished, whereas at the same time the overall efficiency is typically severely reduced. Thus, both linearity and efficiency in back-off require more attention than ever.

This thesis explores various means for enabling the analysis and optimization of the traveling-wave tube for future satellite applications. For this purpose, in Chapter 2, the main components of a traveling-wave tube and their role in the amplification process are briefly described. A qualitative description of the operating principle is given, followed by a brief explanation of the reasons behind the existence of nonlinear distortions in traveling-wave tubes, and how to compensate for them. Important terminology regarding the amplifier characterization is introduced, garnished with an overview over common efficiency definitions.

In Chapter 3, methods and models for the simulation of multi-tone intermodulation products and modulated communication signals are investigated. For this purpose, time-, frequency-, and hybrid-domain codes are considered. Following a short introduction into frequency-domain multi-frequency simulation in MVTRAD, approaches for efficient models for different parts of full-wave helix traveling-wave tube simulation in CST Particle studio are presented. Comparison with both MVTRAD and measurement data is shown. The generated reference CST model is fitted by including cold measurement data and tuning the support rod permittivity values. Continuous-wave interaction simulation results, i.e., gain compression, nonlinear phase shift, and the electron energy spectrum agree well with measurements and an MVTRAD model with directly fitted parameters. Thus, the resulting models enable the study of transients, multi-tone, and multi-carrier simulation, provided that time and computational effort are not crucially limited.

As especially for optimization purposes fast solutions are required, Envelope methods are investigated as hybrid frequency- and time-domain approaches combining advantages of both worlds. They are used to calculate the output behavior of the traveling-wave tube based on input signal quantities and a set of reference data of the traveling-wave tube. Various models from literature are evaluated and adapted for accurate and fast estimation of typical output characteristics. The range of models reaches from simple two-tone approximations for the phase transfer factor and the intermodulation product ratio, over frequency-independent models such as the Static-Gain Curve model, up to the rather involved Quadrature-Polynomial Model.

Results from the validation of MVTRAD, CST Particle Studio, and different Envelope codes are compared, revealing their respective advantages and disadvantages based on the considered scenarios. For continuous-wave and far-spaced multi-tone signals, both CST Particle Studio and MVTRAD have shown to be viable, trustworthy options. But it is also apparent that their not-so-gracious scaling regarding computation time for more intricate settings limits their applicability in many cases, especially considering complex communication signals. Current, typical communication traveling-wave tubes are so flat over frequency, that in typical bands the Static-Gain Curve model suffices for the linearity characterization, especially considering the prediction of powers at the intermodulation levels. The prediction of the phase transfer and faster transients, on the other hand, can benefit from the more sophisticated Quadrature-Polynomial Model. In the case of multi-tone signals with bandwidths, where the Static-Gain Curve model starts to lose accuracy, MVTRAD and CST become more useful. The simulation of future traveling-wave tube types, e.g., folded-waveguides, has shown further areas where the Quadrature-Polynomial Model excels. These delay lines are significantly more dispersive than classical helix lines and bring Envelope models to the limit of their capabilities. Still, the Quadrature-Polynomial Model is able to reliably predict the slope across the output spectrum.

For most simulation and measurement results shown here, a state-of-the-art 150 W commercial reference TWT for communication in Ku-Band is used, such that the extracted statements are most meaningful for practical use. To enhance the occurrence of certain frequency-dependent effects, for some accordingly tagged results, an artificially distorted, more dispersive version of said traveling-wave tube is used in simulation.

In addition to the models themselves, for reproducibly using them, the SIMBA program was written. It incorporates different Envelope methods for the traveling-wave tube with and without linearization in a satellite downlink-chain model with modulation and demodulation of communication signals as defined in the DVB-S2 standard. Various communication characteristics are included to investigate the back-off behavior of the amplifier.

Based on the developed models, together with the experience and data gained throughout their validation, in Chapter 4 the impact of the nonlinearities on typical communication characteristics is evaluated, with the purpose of enhancing the understanding of the relationship between classical and effective modulated characteristics. After a briefly summarized study on the effects on communication signals, simple expressions relating well-known transfer properties from continuous-wave simulation to involved characteristics of multi-tone simulation are derived. Based on a third order intermodulation ratio derivation from literature, a simple expression for the estimation of the noise-power ratio is generated, which offers accurate results up to typical back-off ranges. It has been shown that far in back-off, the noise-power ratio is closely related to the third order intermodulation ratio. As a result, it is possible to translate one to the other by a constant shift in logarithmic scale. Also, to extend the mentioned intermodulation ratio expression from literature, it is combined with a Three-Box Envelope approach.

As the interpretation of the expressions still remains a difficult task, an automated optimization routine based on a previous continuous-wave optimization is developed. For this, an optimizer based on Simulated Annealing is extended to include modulated operation as a part of the goal function. It uses the electron velocity spectrum and noise-power ratio to evaluate the effective collector and beam efficiencies and improve the traveling-wave tube's behavior in modulated back-off.

While this work provides some tools to analyze and optimize the traveling-wave tube for modern communication satellite systems, there are many related topics left to be investigated in future work. For instance, the pool of Envelope model topologies is far from exhausted and element of ongoing research. Also, with DiMoHa, a proprietary, specialized alternative in the time-domain field is under development and has recently shown progress in terms of simulation time and validation. While potentially faster than general purpose tools, its usability and reliability for the simulation of modulated signals is to be evaluated in the future.

Within the scope of this thesis, the collector could only be considered in a reduced manner. Further study on three-dimensional collector simulation with modulated signals could be beneficial for better understanding of the collector's behavior, although the integrated currents have shown to be accurately and consistently predicted. Based on the presented results, a study on collector adaptions for higher collector efficiency in back-off could be of interest. Apart from using more stages to better follow the shape of the electron velocity spectrum, literature offers interesting, more unconventional ideas, e.g., using a time-dynamic, envelope following bias of the collector stages. Such methods might be investigated following the method for calculating the effective electron velocity spectrum. Likewise, this also holds for some other efficiency improvement methods, such as certain waveform-altering signal-injection schemes.

As an alternative to adapting the collector, one could attempt shaping the electron spectrum for better collectability, without sacrifices in terms of beam efficiency and linearity. The many mechanisms leading to a system of trade-offs between most important characteristics so far do not allow for such an ambitious endeavor. Still, complimentary study could bring further light into the darkness of the traveling-wave tube's broadband back-off behavior.

Overall, one can summarize that the modulated operation of traveling-wave tubes offers many challenging research topics. All this shows that even after all these years the traveling-wave tube has not yet reached its limits. It offers plenty of room for improvement to meet the ever rising demands of the satellite communications community and its future applications.

List of Acronyms

APSK	Amplitude- and Phase-Shift Keying
AM-AM	Amplitude-Modulation to Amplitude-Modulation
AM-PM	Amplitude-Modulation to Phase-Modulation
BFM	Bessel-Function Model
CW	Continuous-Wave
DVB-S	*Digital Video Broadcasting – Satellite*
EbNo	Effective Bit-Energy to Noise Power Spectral Density
EVS	Electron-Velocity Spectrum
EVM	Error-Vector Magnitude
FFT	Fast Fourier Transform
FW	Folded-Waveguide
MER	Modulation-Error Ratio
MVTRAD	*MouVemenT RADial*
MVT	MVTRAD
NPR	Noise-Power Ratio
PIC	Particle-In-Cell
PAPR	Peak-to-Average Power Ratio
PPM	Periodic Permanent Magnet
PSK	Phase-Shift Keying
PSB	Poza-Sarkozy-Berger
QPSK	Quadrature Phase-Shift Keying
QPM	Quadrature Polynomial Model
QAM	Quadrature-Amplitude Modulation
RF	Radio-Frequency
RRC	Root-Raised Cosine
SNR	Signal-to-Noise Ratio
SA	Simulated Annealing
SIMBA	*Simulation of Inter-Modulations in Back-Off and Applications*
SSPA	Solid-State Power Amplifier
SGC	Static-Gain Curve
TBM	Three-Box Model
TWT	Traveling-Wave Tube

List of Figures

List of Tables

Bibliography

[1] A. Gilmour, *Principles of Traveling Wave Tubes*, ser. Artech House radar library. Artech House, 1994.

[2] R. J. Barker, N. C. Luhmann, J. H. Booske, and G. S. Nusinovich, *Modern Microwave and Millimeter-Wave Power Electronics*. IEEE Press, Apr. 2005.

[3] J. A. Eichmeier and M. Thumm, *Vacuum Electronics: Components and Devices*. Springer Berlin Heidelberg New York, 2008.

[4] G. Kornfeld and E. Bosch, “From History to Future of Satellite TWT Amplifiers”, *Frequenz*, vol. 55, no. 9-10, pp. 258–262, 2001.

[5] J. M. Weekley and B. J. Mangus, “TWTA versus SSPA: A Comparison of On-Orbit Reliability Data”, *IEEE Transactions on Electron Devices*, vol. 52, no. 5, pp. 650–652, May 2005.

[6] E. F. Nicol, B. J. Mangus, J. R. Grebliunas, K. Woolrich, and J. R. Schirmer, “TWTA Versus SSPA: A Comparison Update of the Boeing Satellite Fleet On-orbit Reliability”, in *14th IEEE International Vacuum Electronics Conference (IVEC)*, May 2013, pp. 1–2.

[7] E. F. Nicol and J. M. Robison, “TWTA On-Orbit Reliability for Satellite Industry”, *IEEE Transactions on Electron Devices*, vol. 65, no. 6, pp. 2366–2370, Jun. 2018.

[8] European Telecommunications Standards Institute, “Digital Video Broadcasting (DVB); Second Generation Framing Structure, Channel Coding and Modulation Systems for Broadcasting, Interactive Services, News Gathering and other Broadband Satellite Applications”, en, European Telecommunications Standards Institute, Standard ETSI EN 302 307-1, 2014.

[9] ——, “Digital Video Broadcasting (DVB); Second Generation Framing Structure, Channel Coding and Modulation Systems for Broadcasting, Interactive Services, News Gathering and other Broadband Satellite Applications; Part 2: DVB-S2 Extensions (DVB-S2X)”, en, European Telecommunications Standards Institute, Standard ETSI EN 302 307-2, 2015.

[10] D. S. Komm, R. T. Benton, H. C. Limburg, W. L. Menninger, and X. Zhai, “Advances in Space TWT Efficiencies”, *IEEE Transactions on Electron Devices*, vol. 48, no. 1, pp. 174–176, Jan. 2001.

[11] P. Waller, “Modélisation numérique de l’interaction et diagnostic expérimental du faisceau d’électrons dans un tube à ondes progressives spatial”, PhD thesis, Université Paris 7, Paris, 1999.

[12] T. M. Antonsen Jr and B. Levush, “CHRISTINE: A Multifrequency Parametric Simulation Code for Traveling Wave Tube Amplifiers”, Naval Research Lab Washington DC Vacuum Electronics Branch, Tech. Rep., 1997.

[13] *CST Particle Studio*, `https://www.cst.com/`.

[14] *MAGIC Tool Suite*, `https://www.orbitalatk.com/magic/`.

[15] S. Meyne, “Simulation and Design of Traveling-Wave Tubes with Folded-Waveguide Delay Lines”, PhD thesis, Technische Universität Hamburg-Harburg, 2017.

[16] R. E. Collin, *Foundations for Microwave Engineering*, ser. IEEE Press on Electromagnetic Wave Theory. John Wiley & Sons., 2001.

[17] J. R. Pierce, “Traveling-Wave Tubes”, *The Bell System Technical Journal*, vol. 29, no. 2, pp. 189–250, 1950.

[18] J. Copeland and A. A. Haeff, “Andrei Haeff and the Amazing Microwave Amplifier”, *IEEE Spectrum*, vol. 2015, no. 9, pp. 32–37, 2015.

[19] A. Katz, “Linearization: Reducing distortion in power amplifiers”, *IEEE Microwave Magazine*, vol. 2, no. 4, pp. 37–49, 2001.

[20] P. Thouvenin, D. Henry, and A. Pelletier, “New Helix Tapers Boost Space TWT Efficiency to 55%, Broadband”, in *1987 International Electron Devices Meeting*, doi: `10.1109/IEDM.1987.191463`, vol. 33, 1987, pp. 477–480.

[21] F. André, P. Bernardi, N. M. Ryskin, F. Doveil, and Y. Elskens, “Hamiltonian Description of Self-Consistent Wave-Particle Dynamics in a Periodic Structure”, *EPL (Europhysics Letters)*, vol. 103, no. 2, 2013, doi: `10.1209/0295-5075/103/28004`.

[22] D. F. G. Minenna, Y. Elskens, F. André, A. Poyé, J. Puech, and F. Doveil, “Dimoha: A time-domain algorithm for traveling-wave tube simulations”, *IEEE Transactions on Electron Devices*, vol. 66, no. 9, pp. 4042–4047, Sep. 2019.

[23] S. Sensiper, “Electromagnetic Wave Propagation on Helical Conductors”, *Massachusetts Institute of Technology Technical Report*, 1951.

[24] P. K. Tien, “Traveling-Wave Tube Helix Impedance”, *Proceedings of the IRE*, vol. 41, no. 11, pp. 1617–1623, Nov. 1953.

[25] F. Shen, Y. Wei, H. Yin, Y. Gong, X. Xu, S. Wang, W. Wang, and J. Feng, “A Novel V-Shaped Microstrip Meander-Line Slow-Wave Structure for W-band MMPM”, *IEEE Trans. on Plasma Science*, vol. 40, no. 2, pp. 463–469, Feb. 2012.

[26] G. Ulisse and V. Krozer, “W-Band Traveling Wave Tube Amplifier Based on Planar Slow Wave Structure”, *IEEE Electron Device Letters*, vol. 38, no. 1, pp. 126–129, Jan. 2017.

[27] X. Bian, M. Miao, and Z. Li, “Design and Analysis of Micro-Machined Folded Waveguide Slow-Wave Structure for 220 GHz Application”, in *2016 IEEE MTT-S Int. Microwave Workshop Series on Advanced Materials and Processes for RF and THz Applications (IMWS-AMP)*, doi: `10.1109/IMWS-AMP.2016.7588384`, 2016.

[28] F. Li, T. Ma, and Y. Liu, “Design of W Band Three-Section Folded Waveguide Traveling Wave Tube”, in *2016 IEEE Int. Vacuum Electronics Conf. (IVEC)*, doi: `10.1109/IVEC.2016.7561897`, Apr. 2016, pp. 1–2.

[29] D. F. G. Minenna, Y. Elskens, F. André, J. Puech, A. Poyé, F. Doveil, and T. Pereira, “DIMOHA: Traveling-Wave Tube Simulations Including Band Edge and Multiple-Carriers Operations”, in *2019 International Vacuum Electronics Conference (IVEC)*, Apr. 2019, pp. 1–2.

[30] D. Safi, P. Birtel, S. Meyne, and A. F. Jacob, “Helix Traveling-Wave Tube Interaction Simulation with CST Particle Studio”, in *2017 IEEE Int. Vacuum Electronics Conf. (IVEC)*, 2017.

[31] ——, “A Traveling-Wave Tube Simulation Approach with CST Particle Studio”, *IEEE Transactions on Electron Devices*, vol. 65, no. 6, pp. 2257–2263, 2018.

[32] I. Langmuir and K. B. Blodgett, “Currents Limited by Space Charge Between Concentric Spheres”, *Physical Review*, vol. 24, no. 1, 1924, doi: `10.1103/physrev.24.49`.

[33] J. E. Rowe, *Nonlinear Electron-Wave Interaction Phenomena*. Academic Press, 2013, doi: `10.1016/c2013-0-08102-9`.

[34] P. R. R. Rao and S. K. Datta, “Estimation of Conductivity Losses in a Helix Slow-Wave Structure Using Eigen-Mode Solutions”, in *2008 IEEE International Vacuum Electronics Conference*, doi: `10.1109/IVELEC.2008.4556452`, Apr. 2008, pp. 99–100.

[35] C. L. Kory and J. A. Dayton, “Investigation of the effect of support rod permittivity on TWT performance”, in *IEEE Conference Record - Abstracts. 1997 IEEE International Conference on Plasma Science*, May 1997, p. 199.

[36] R. M. Gagliardi, *Introduction to Communications Engineering*, ser. A Wiley-Interscience Publication. Wiley, 1978.

[37] J. Bretting and P. Treytl, “Größen zur Charakterisierung der Nichtlinearitäten von Mikrowellen-Sendeverstärkern und deren Zusammenhänge”, *NTZ-Kurier, Berlin*, 1972.

[38] J. Bretting, “Der Betriebswirkungsgrad von Wanderfeldröhren unter Berücksichtigung von nichtlinearen Verzerrungen”, PhD thesis, RWTH Aachen, 1971.

[39] G. Gatti, W. Bosch, S. C. Soriano, R. M. R. de Gopegui, and Y. Vichery, “IMAL-2 a Simulation Software for Inter-Modulation Analysis of Power Amplifiers”, *XRM Tech. Notes*, 1996.

[40] D. Gabor, “Theory of Communication. Part 1: The Analysis of Information”, *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, vol. 93, no. 26, pp. 429–441, 1946.

[41] B. Boashash and G. Jones, *Instantaneous Frequency and Time-Frequency Distributions*. Longman Cheshire, 1992.

[42] L. Marple, “Computing the Discrete-Time "Analytic" Signal via FFT”, *IEEE Transactions on Signal Processing*, vol. 47, no. 9, pp. 2600–2603, Sep. 1999.

[43] D. Safi, P. Birtel, S. Meyne, and A. F. Jacob, “Comparative Study of Envelope Models for Multi-Tone Simulation of Traveling-Wave Tubes”, in *2017 47th European Microwave Conference (EuMC)*, 2017, pp. 852–855.

[44] D. Safi, P. Birtel, S. Meyne, and A. F. Jacob, “Comparative Study of Envelope Models for Hybrid Time- and Frequency-Domain Simulation of Traveling-Wave Tubes”, *International Journal of Microwave and Wireless Technologies*, vol. 10, no. 5-6, pp. 521–529, 2018.

[45] P. J. Loughlin and B. Tacer, “Comments on the Interpretation of Instantaneous Frequency”, *IEEE Signal Processing Letters*, vol. 4, no. 5, pp. 123–125, 1997.

[46] L. Mandel, “Interpretation of Instantaneous Frequencies”, *American Journal of Physics*, vol. 42, no. 10, pp. 840–846, 1974.

[47] G. Strang, M. Krieger, and K. Lippert, *Wissenschaftliches Rechnen*, ser. Springer-Lehrbuch Masterclass Series. Springer Berlin Heidelberg, 2010.

[48] M. Knorrenschild, *Numerische Mathematik: Eine beispielorientierte Einführung*. Carl Hanser Verlag GmbH & Company KG, 2017.

[49] F. N. Fritsch and R. E. Carlson, "Monotone Piecewise Cubic Interpolation", *SIAM Journal on Numerical Analysis*, vol. 17, no. 2, pp. 238–246, 1980.

[50] A. A. Saleh, "Frequency-Independent and Frequency-Dependent Nonlinear Models of TWT Amplifiers", *IEEE Transactions on Communications*, vol. 29, no. 11, pp. 1715–1720, 1981.

[51] M. T. Abuelmaatti, "Frequency-Dependent Nonlinear Quadrature Model for TWT Amplifiers", *IEEE transactions on communications*, vol. 32, pp. 982–986, 1984.

[52] H. B. Poza, Z. A. Sarkozy, and H. L. Berger, "A Wideband Data Link Computer Simulation Model", in *NAECON'75; Proceedings of the National Aerospace and Electronics Conference*, 1975, pp. 71–78.

[53] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems*. New York, NY, USA: Plenum Press, 1992.

[54] L. Buderath, "Investigation of Nonlinear Multi-Box Models for Time-Domain Simulation of Traveling-Wave Tubes", Bachelor's Thesis, TU Hamburg, Mar. 2017.

[55] P. Hetrakul and D. P. Taylor, "Nonlinear Quadrature Model for a Travelling-Wave-Tube-Type Amplifier", *Electronics letters*, vol. 11, no. 2, pp. 50–50, 1975.

[56] E. H. Moore, "On the Reciprocal of the General Algebraic Matrix", *Bull. Am. Math. Soc.*, vol. 26, pp. 394–395, 1920.

[57] R. Penrose, "A Generalized Inverse for Matrices", in *Mathematical Proceedings of the Cambridge Philosophical Society*, Cambridge University Press, vol. 51, 1955, pp. 406–413.

[58] C. R. Johnson, *Matrix Theory and Applications*. American Mathematical Soc., 1990, vol. 40.

[59] L. R. Turner, "Inverse of the Vandermonde Matrix with Applications", *NASA Technical Note*, 1966.

[60] R. W. Koch, "Random Signal Method of Nonlinear Amplitude Distortion Measurement", *IEEE Transactions on Instrumentation and Measurement*, vol. IM-20, no. 2, pp. 95–99, May 1971.

[61] B. Evans, *Satellite Communication Systems*, ser. IET Telecommunications. Institution of Engineering & Technology, 2011.

[62] S. O. Wallander, "Reflexions and Gain Ripple in TWT's", *IEEE Transactions on Electron Devices*, vol. 19, no. 5, pp. 655–660, 1972.

[63] S. Meyne, P. Birtel, and A. F. Jacob, "Q-Band High-Power Traveling-Wave Tube with Optimized Folded-Waveguide Delay Line", in *2015 IEEE Int. Vacuum Electronics Conf. (IVEC)*, doi: `10.1109/IVEC.2015.7223780`, Apr. 2015, pp. 1–2.

[64] S. Meyne, P. Bernadi, P. Birtel, J. F. David, and A. F. Jacob, "Large-Signal 2.5-D Steady-State Beam-Wave Interaction Simulation of Folded-Waveguide Traveling-Wave Tubes", *IEEE Trans. on Electron Devices*, vol. 63, no. 12, pp. 4961–4967, Dec. 2016, doi: `10.1109/TED.2016.2616948`.

[65] J. Boyhan, "Normalized Power Spectra for FDM/FM Signals and Noise Power Ratio Calculations", *IEEE Transactions on Communications*, vol. 32, no. 1, pp. 77–81, Jan. 1984.

[66] A. Darbandi and W. Zoccarato, "High Efficiency, Linearized Power Amplifier for Communication Satellites", in *1992 22nd European Microwave Conference*, vol. 2, Sep. 1992, pp. 1131–1136.

[67] C. Devieux, "Nonlinear Spectrum-Spreading Interference Effects Between Adjacent Satellite Transponders", *IEEE Transactions on Communications*, vol. 23, no. 11, pp. 1379–1385, Nov. 1975.

[68] O. Nilsson, *Nonlinear Distortion In Traveling Wave Tubes*, ser. Chalmers Univ of Technology Goeteborg, Research Report (chalmers Tekniska Högskola. Forskningslaboratorium Foer Elektronik). Defense Technical Information Center, 1966.

[69] J. D. Wilson, "A Simulated Annealing Algorithm for Optimizing RF Power Efficiency in Coupled-Cavity Traveling-Wave Tubes", *IEEE Transactions on Electron Devices*, vol. 44, no. 12, pp. 2295–2299, Dec. 1997.

[70] D. R. Whaley, L. K. Behnke, and J. A. McKay, "Application-Specific Multi-Parameter TWT Optimization", in *2007 IEEE International Vacuum Electronics Conference*, May 2007, pp. 1–2.

[71] J. A. David, C. L. Kory, H. T. Tran, R. L. Ives, and D. Chernin, "Enhanced Features for Design of Traveling Wave Tubes Using CHRISTINE-1D", *IEEE Transactions on Plasma Science*, vol. 35, no. 4, pp. 1056–1064, Aug. 2007.

[72] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, *et al.*, "Optimization by Simulated Annealing", *science*, vol. 220, no. 4598, pp. 671–680, 1983.

[73] J. X. Qiu, D. K. Abe, T. M. Antonsen, B. G. Danly, and B. Levush, “Traveling-Wave Tube Amplifier Performance Evaluation and Design Optimization for Applications in Digital Communications with Multilevel Modulations”, *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 8, pp. 1911–1919, Aug. 2003.

[74] D. Safi, P. Birtel, M. Wulff, S. Meyne, and A. F. Jacob, “Back-off Efficiency Optimization Of Traveling-wave Tubes Using Simulated Annealing”, in *2018 IEEE International Vacuum Electronics Conference (IVEC)*, Apr. 2018, pp. 169–170.

[75] S. Meyne, M. Krenz, and A. F. Jacob, “Taper Optimization for Helix Traveling-Wave Tubes Using Adaptive Simulated Annealing”, in *IEEE International Vacuum Electronics Conference (IVEC)*, Apr. 2014, pp. 143–144.