Chapter 2
Theoretical Background of Thermal Transient Measurements

Gábor Farkas, András Poppe, and Márta Rencz

In this chapter we collected the most important background knowledge that is needed to understand thermal transient measurements.

Speaking about measurements, we need to remember that a measurement is always accompanied by an inherent modeling step. Measuring the size of an object and claiming its length, width, and height is equivalent to replacing it with a model, which is a single (rectangular) block, and describing this model by these three quantitative parameters.

In thermal analysis the modeling of the system is much more challenging. All the physical quantities that play a role in thermal measurements must be precisely defined to avoid ambiguity.

2.1 Temperature and Heat Transfer

Temperature is the manifestation of the thermal energy of a finite size object. Thermal energy is the internal energy associated with the stochastic movement of particles in the object. These particles can be molecules in a fluid or gas, crystal lattice atoms in solids, or electrons in an electrically conductive material.

Heat is the internal energy, which is transferred between two or more finite size objects by various mechanisms at the level of particles (atoms and molecules). Heat transfer can be accompanied but does not need to involve transfer of matter.

G. Farkas
Siemens Digital Industry Software STS, Budapest, Hungary

A. Poppe (✉) · M. Rencz
Siemens Digital Industry Software STS, Budapest, Hungary

Budapest University of Technology and Economics, Budapest, Hungary
e-mail: Poppe.Andras@vik.bme.hu

M. Rencz et al. (eds.), *Theory and Practice of Thermal Transient Testing of Electronic Components*, https://doi.org/10.1007/978-3-030-86174-2_2

The quantity of energy transferred as heat can be measured by its effect on the states of interacting bodies. Such effects can include the amount of matter participating in a phase change (e.g., amount of ice melted) or the change in temperature of a body dedicated to measuring the amount of transferred energy (temperature sensor, thermometer).

Power is the rate per unit time at which energy is applied externally on the system, or transferred between portions of the system.

Heat flux is the intensity of the transfer of energy per unit area per unit of time, that is, the power applied or forced through a unit of area.

The conventional symbol used to represent the temperature is T; the amount of heat transferred is denoted by Q. The SI unit of temperature is kelvin (K) or centigrade (°C); the unit of heat (also known as *thermal energy*) is joule (J, Ws).

In this work power is denoted by P and heat flux by φ. The SI unit of power is watt (W); the unit of heat flux is watt per square meter (W/m^2).

The primary mechanism of thermal energy transfer in electronic systems, as they are mostly solids, is *conduction*, at direct contact of objects, within molecular dimensions. The transfer occurs by the stochastic motion of particles, which can be "electrons" and "phonons" where the latter is the quantized lattice vibration.

Convection is a heat transfer mechanism in which one body heats another over macroscopic distances, through an intermediate circulating medium that carries energy from a boundary of one to a boundary of the other. The heat transfer on the surface of respective solid bodies towards the medium occurs by conduction.

In the related discipline of physics, in *fluid mechanics*, all media such as fluids or gases, aerosols, etc. are denoted as a generalized "fluid." Convection always involves the motion of matter. The internal energy of the medium is influenced not only by the stochastic motion of particles; it can be changed directly by *thermodynamic work*, by mechanisms that act macroscopically on the system, for example, by the motion of a piston.

Radiation is a heat transfer mechanism that occurs between separated or even remote bodies by means of electromagnetic waves. Accordingly, it requires no medium; it transfers heat over transparent matter or vacuum. All solid bodies emit radiation because of the stochastic motion of charged particles; this radiation grows in a "temperature to the fourth" manner.

The direction of heat transfer is always from the hotter to the cooler matter portions, as long as the temperature difference exists between them.

The heat transfer in solids is governed by local thermal properties: thermal conductivity and specific heat. These thermal parameters are temperature dependent in the semiconductor and package materials, which are most frequently used in electronics. However, the change of these parameters in the temperature range of the typical use is rather flat. This means that in many practical cases, the material parameters can be considered temperature independent, which simplifies the calculations, allowing in many cases the use of linear relationships.

Several thermal interface materials, such as thermal pastes and sheets, are anisotropic; they perform differently in different directions. Still, their orientation does not change during their operation.

In some thermal interface materials also phase change can occur at higher temperatures. Phase change accumulates or releases a large amount of heat, for example, phase change mechanisms enable intensive heat transfer in heat pipes.

2.2 Thermal Equilibrium, Steady State, and Thermal Transients

According to definition, two physical systems are in *thermal equilibrium* if there is no net flow of thermal energy between them when they are connected by a path permeable to heat. Extending this definition to *different portions of a system*, we consider a system being in thermal equilibrium when all parts of the system are at the same temperature. When one of the "systems" is the outer world, then the equilibrium is reached at the temperature of the ambient.

Steady state means that the temperature in different portions of the system does not change with time. Nonequilibrium states can be steady states if there is a source of energy to maintain the nonequilibrium condition. Without the source of energy, the system would settle into an equilibrium state after a certain time.

Thermal transient is a process through which the system or its portions transit from one temperature to another temperature.

In a heating process, the system moves from a lower temperature state to a higher one, and in a cooling process, the system starts from a higher and arrives at a lower temperature state.

Heating transients are always the result of adding energy to the system. This energy surplus is often applied on thermal systems as a time-dependent $P(t)$ power profile at one or more entry points, "driving points" over a time interval.

In electronic systems the energy that heats the system is in most cases the introduced electrical energy.

In cooling processes the energy leaves the system in the form of dissipation, that is, in the form of heat. Cooling can be a relaxation after revoking all power from the entry points for a prolonged time, or returning to a lower energy state at diminished power level.

The word "dissipation"[1] is used in a loose interpretation in the technical literature.

If the energy entry occurs at an area which is small compared to the size of the system, then that location is frequently called *junction*. The term means in thermal engineering a spot, which is considered to be isothermal and emitting homogeneous heat flux.

This name is inherited from power electronics where the heat source is in many cases a thin "dissipating," more precisely *heat-generating* layer near the upper

[1] In its original grammatical sense, "to dissipate" means *to scatter*, *to throw away*, *to dispose of* something.

surface of a semiconductor device. In many active power devices, this layer is in fact a *pn* junction, an area where semiconductors of different types join each other.

Proper distinction between different interpretations of this term will be made case by case in the subsequent chapters.

The development of the temperature change in time is highly influenced by the internal geometry and material properties of the system. Consequently, a systematic analysis of the transient process can yield relevant information on the structural composition of the system.

In the characterization of systems, those special transients play an eminent role where the system transits from one steady state to another steady state. Such "finished" transients yield the most complete information on the internal structure of the system. The structural details are determined with the best resolution in the vicinity of the power entry. From shorter transients where the steady state is not yet reached in all parts of the system, only limited information can be gained on the entire system structure.

2.3 Thermal Processes and Their Modeling in Electronic Systems

In electronic systems the most relevant heat sources are semiconductor devices; therefore, their junction temperature is a critical parameter influencing the reliability and the lifetime of the system.

The power generated in these devices can be calculated from the voltage and current values at their pins. These values change in time in a complex way and are determined by the electric characteristics of the devices, which are highly nonlinear, and *temperature dependent*. Typical device characteristics and their temperature dependence are discussed in detail in Chap. 6.

In many cases an electronic equipment operates in a relatively narrow temperature range, such as 0 °C–120 °C in laboratories, which is 273 K–393 K on the absolute temperature scale. In this range the temperature dependence of the thermal conductivity and specific heat of the used materials is usually negligible. In a system composed of materials of temperature-independent thermal properties, the flow of thermal energy is proportional to the temperature *differences*; the actual absolute temperature level has only minor influence.

For this reason, as long as the heat propagation takes the form of heat conduction and convection, the thermal behavior of electronic systems composed of semiconductor chips, their packages, and cooling mounts can be well described with the mathematical apparatus of the *theory of linear systems*. The linear approach can be used only with severe limitations in the case of systems with phase change materials, or systems, which operate at elevated temperatures where radiation from the outer surfaces becomes significant in the heat removal process.

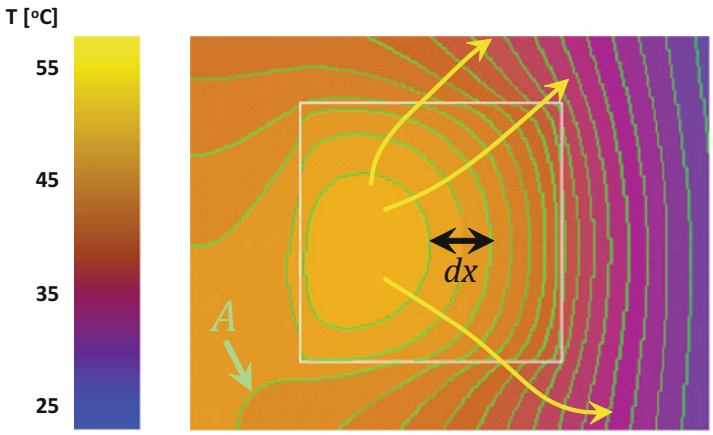

Fig. 2.1 Cross-sectional view of three-dimensional conductive heat transfer in solid material. Isotherms and heat-flow trajectories calculated in a thermal simulator are shown

The investigation of transient processes needs appropriate models of the thermal systems. These models can be of *continuous* or *distributed* nature, based on the differential equations governing the heat conduction in solids and the convection in gases and fluids. Such *detailed*, *continuous* models are usually analyzed numerically by finite element or finite difference software tools.

For analyzing all heat transfer mechanisms defined above, *computational fluid dynamics* (CFD) solvers, such as [56], offer a tool to simulate the conjugate heat and mass transfer.

The software tools display the simulation results in various forms. *Trajectories* demonstrate how the heat spreads in conductive regions and how the "fluid" flows in convective ones; *isothermal surfaces* denote where the temperature is equal at a certain time moment. It has to be noted that the geometric boundaries of physical objects and interface layers, which are often planar and rectangular in a real system, rarely coincide with isothermal surfaces, and the shape of the latter dynamically changes during a transient process.

Figure 2.1 visualizes the conductive heat transfer in a solid body. The different colors correspond to the temperature in the material, isothermal surfaces follow each other with an equidistant temperature difference, and the trajectories of the heat spreading are represented by curved arrows that are perpendicular to the isothermal surfaces.

The reference above to "finite element" or "finite difference" tools indicates that even the analytic equations that describe continuous (distributed) systems must be converted to their numerical counterparts, that is, a continuous system description must be *discretized*. This means that the continuously distributed material is lumped into small pieces for being analyzed by numerical methods in a computer. A net of adjoining lumps generated in the discretizing process of the model is often called a *mesh*. The lumps convey energy into each other in case of conduction and convey energy and matter into each other in case of convection. The abstract representation

of the mesh that constitutes the numerical model of the distributed system is a set of *nodes* and *links* between the nodes. The graph of these nodes and links is called a *network* in the world of electrical engineering. This way, many techniques used in *network theory* are borrowed for the analysis of heat conduction problems in solids.

2.3.1 Equivalent Linear Models

The supposed linearity in the thermal domain implies that when increasing the power levels in the system, the temperature grows nearly proportionally. This assumption does not apply to cases with considerable nonlinear effects, e.g., systems with turbulent flow or significant radiation, or cases where the thermal conductivity of semiconductors exhibits strong temperature dependence, but these phenomena are usually negligible in the temperature range specified above.

This book investigates *time-invariant* thermal systems, in which the geometrical structure and material properties do not vary in time. Certain thermal interface types such as pastes tend to change their thickness during use, especially when pressure and power are applied on them the first time; successive transients yield slightly different results. Similarly, adhesives change their composition at initial curing or at their first use. Such effects are discussed in [63].

The effects of *wear* and *degradation* may cause significant alteration in the shape and thermal properties of some system components. These effects are treated in depth in the literature; some aspects that are closely related to the concept of thermal transient testing are discussed in Sect. 7.4. Still, the initial changes in system composition or later degradation can be observed rarely in the time span of a single transient; it is appropriate to use the time-invariant approach.

Linear time-invariant (LTI) systems can be analyzed in both continuous and discretized approaches. The discretization of a continuous body through a rectangular mesh is illustrated in Fig. 2.2.

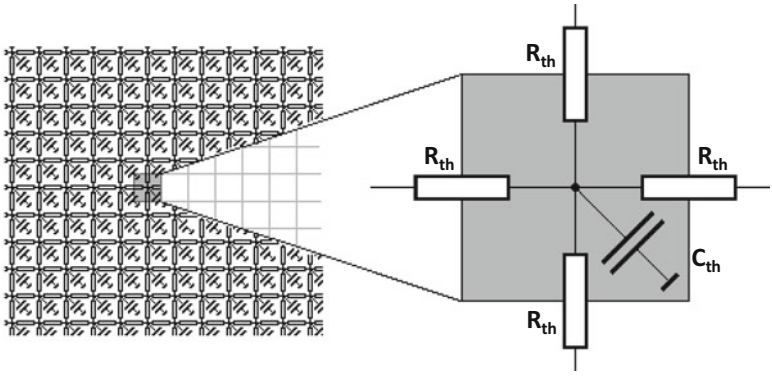

Fig. 2.2 Resistor-capacitor (RC) network model of a real physical structure

Taking the primary effect that is heat conduction, we experience a blend of two energy exchange mechanisms, heat propagation through an elementary portion of material and internal energy growth due to heat flow into that portion. The two mechanisms occur simultaneously in all system portions located anywhere in *space* and at any moment of *time*.

For simplicity, let us analyze first the two mechanisms separately. The energy storage and the temperature change can be formulated in an integral form on larger system regions and in a differential form for elementary portions.

In a *larger region* which contains no heat sources, the Q thermal energy grows *in time* as a result of φ heat flux flowing through all segments of its A surface. The heat flux integrated over the surface yields the P power applied on the region.

A changing $P(t)$ power causes a $Q(t)$ growth of energy in the V volume and m mass of the region. This energy growth manifests as $T(t)$ temperature elevation. The proportionality between thermal energy change and temperature change is expressed as C_{th} thermal capacitance.

Figure 2.1 demonstrates a case when an internal domain contains a heat source. Again, the P power can be interpreted as the sum of all heat trajectories, which intersect the A surface of a region enclosed by an isothermal boundary. The power is of the same P value on each A surface of isothermal shells containing the same heat source.

A *temperature difference between* two surface segments of a larger region induces a heat flow from the hotter towards the cooler one. Keeping this difference for a prolonged time steady state is reached; the heat flow stabilizes. The sum of all heat flux trajectories in a cross section of the region is now a steady P_{const} power, and each slice along the heat flow injects this power into the next region. The proportionality between the power and the temperature difference is expressed as the R_{th} thermal resistance.

The detailed temperature distribution and its change in time can be explored considering elementary portions. Suppose the elementary portion is cut out of the material between isothermal surfaces; it is of a small surface along which the thickness is of constant dx infinitesimal value.

Some parts in assemblies of power electronics correspond to a sandwich-like structure, in which all heat flows in a dedicated x direction. If uniform power is applied on a layer of such a laminate, a homogeneous heat flux will flow through all different material layers. In the following, the surface over which a homogeneous flux flows will be uniformly denoted by A, whether an elementary section or a whole laminate layer is considered, because the related equations are of the same format.

Figure 2.3 illustrates how the difference of the heat flux entering and leaving an elementary portion increases the thermal energy in it. The flux difference causes dQ energy growth and dT temperature growth in a dt time instant:

$$\Delta P = (\varphi_{\text{in}} - \varphi_{\text{out}}) \cdot A = dQ/dt = -c_{\text{p}} \cdot m \cdot dT/dt \tag{2.1}$$

In (2.1) c_{p} denotes the specific heat of the material, and m is the mass of the section. The φ_{out} flux is forwarded into the next portion of matter, which behaves again in the same way.

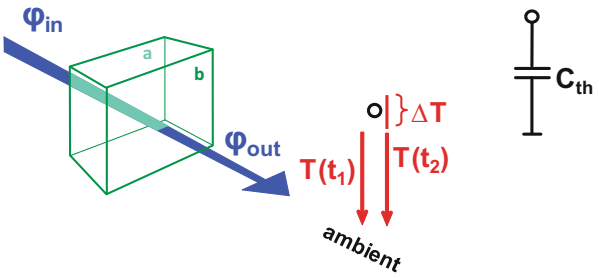

Fig. 2.3 Heat flow into a material slice

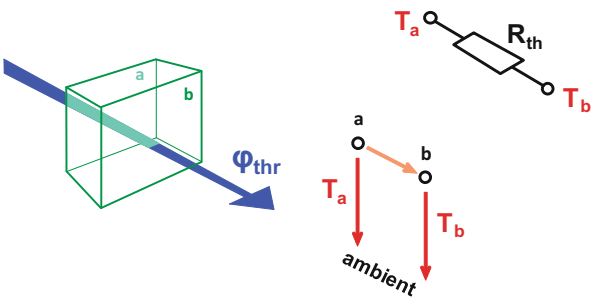

Fig. 2.4 Heat flow through a material slice

It is easy to measure the geometric size of an actual structural element, in this case the A surface and dx length. On the other hand, determining the mass of a segment of the structure is rarely feasible. For this reason it is more practical to introduce another material property, the c_V volumetric specific heat, also called volumetric thermal capacitance into (2.1):

$$\Delta P = dQ/dt = -c_V \cdot V \cdot dT/dt \tag{2.2}$$

$V = dx \cdot A$ is the volume of the slice; $c_V = c_p \cdot \rho$, and ρ is the material density.

Investigating the heat flow *through* the portion (Fig. 2.4), we find that in the continuous approach, it obeys the differential form of Fourier's law:

$$\varphi_{thr} = -\lambda \cdot \nabla T \tag{2.3}$$

where φ_{thr} is the heat flux across the section boundary, λ is the thermal conductivity in that section of material and ∇T is the gradient (spatial derivative) of the temperature.[2]

Equation (2.3) expresses that φ is a vector and the heat flows towards the cooler portions of a body.

[2] In the literature the thermal conductivity is denoted by k or λ. In this book we use λ; we reserve k for denoting the Boltzmann constant. The heat flux is often denoted as q, but in this book q is used for the elementary charge.

In a dx infinitesimal slice of A surface, (2.3) takes the form:

$$P = \varphi_{\text{thr}} \cdot A = -\lambda \cdot A \cdot dT/dx. \tag{2.4}$$

It can be observed in the figures and in the related equations that the dT quantity denotes in (2.4) a spatial temperature difference and in (2.1) a temporal difference. In (2.1) ΔP is a difference in space, the power difference between the two surfaces where the heat flow enters and leaves.

In electronics the heat dissipation is localized to the small volumes described as "junction" above. In other regions, the energy growth in a slice is the result of the trapped heat flux. Hence, the previous considerations can be combined into the classical heat equation, which is expressed in one dimension as

$$\frac{dT}{dt} = \left(\frac{\lambda}{c_{\text{V}}}\right) \cdot \frac{d^2T}{dx^2} = \alpha \cdot \frac{d^2T}{dx^2}. \tag{2.5}$$

The thermal diffusivity α, defined as $\alpha = \lambda/c_{\text{V}}$, is the measure of thermal inertia. In a material of high thermal diffusivity, heat moves rapidly; the substance conducts heat quickly relative to its volumetric heat capacity.

Discretizing the structure into small lumps of homogeneous material condenses (2.4) into thermal resistances. A lump has now dx length along the heat flow and A surface perpendicular to it; and a temperature drop of $T_{\text{a}} - T_{\text{b}}$ occurs between its a and b isothermal faces (Fig. 2.4). Now (2.4) takes the form:

$$T_{\text{a}} - T_{\text{b}} = P \cdot R_{\text{th}} = P\left(\frac{1}{\lambda}\frac{dx}{A}\right), \qquad R_{\text{th}} = \left(\frac{1}{\lambda}\frac{dx}{A}\right) \tag{2.6}$$

In the discretized representation, the two faces of the section are connected by an R_{th} thermal resistance.

Similarly, in a material lump exposed to continuous φ heat flux over A area resulting in ΔP power difference between the entry and the exit, the energy change in a short $dt = t_2 - t_1$ time interval is from (2.1)

$$dQ = \Delta P \cdot dt = C_{\text{th}} \cdot (T_2 - T_1), \tag{2.7}$$

where $T_1 = T(t_1)$ is the temperature of the material at t_1 time and $T_2 = T(t_2)$ is the temperature of the material at t_2 time.

C_{th} is the thermal capacitance of the slice:

$$C_{\text{th}} = c \cdot m = c \cdot \rho \cdot dx \cdot A, \quad C_{\text{th}} = c_{\text{V}} \cdot V = c_{\text{V}} \cdot dx \cdot A. \tag{2.8}$$

In the discretized model of a complete system as represented in Fig. 2.2, also the topology in which the R_{th} and C_{th} constituents are connected is to be defined.

Temperature is a quantity, which is measured related to a reference value. When it is expressed, for example, in centigrade, then the temperature is related to the reference level of the internal energy of melting ice. In (2.6) the T_a and T_b temperatures are measured with respect to the reference level, and so are T_1 and T_2 in (2.7). Still, because of the linear approach, regardless of which reference level was chosen, it disappears from the equations when the temperatures are subtracted.

In thermal transient measurements, it is generally assumed that the near environment of the measurement arrangement such as the air temperature in the laboratory or the temperature of the circulated coolant in a cold plate device does not change during the measurement time. Similarly, CFD simulation always aims at analyzing a limited part of the universe only; the external world is often represented as a constant temperature on the system boundary. This constant temperature attributed to the near environment is called *ambient temperature* or *ambient* in short, usually denoted by T_A (or T_{amb}) in the literature. The ambient is the thermal counterpart of the electrical ground (zero reference potential, *datum reference* in other engineering disciplines).

The electrical networks are extremely abstract; a lump with its "volume" and "faces" appears as a node in the graph of a circuit scheme. The reference level in electronics is named "ground" and is represented by a $\perp$ sign.

Thus, as already shown in Fig. 2.2, the C_{th} thermal capacitance appears between the node representing the material portion and the ambient. The R_{th} thermal resistance connects two such nodes; the reference towards the ground disappears.

Equations (2.3)–(2.8) are of the same format as the descriptive differential equations of electronics, replacing the P power by the I electrical current, the T temperature by the U electrical potential, and the ΔT temperature difference by the V electrical voltage.[3] The thermal resistances and capacitances correspond to their electric counterparts of similar name.

In an electrical network, the current flows through the net of resistances and capacitances and causes a voltage drop (potential difference) between the nodes of the circuit. The current is measured in amperes and the voltage in volts; resistances and capacitances are measured in ohms (V/A) and farads (As/V), respectively.

In the thermal network, heat, that is power, flows through the net of thermal resistances and capacitances and causes a temperature drop between the nodes of the network. The power is measured in watts and the temperature in kelvins (K) or centigrade (°C); thermal resistances and capacitances are measured in K/W and Ws/K (J/K), respectively.

In network theory those networks in which excitation (powering) occurs at a single specific point can be characterized by their *thermal impedance*. This concept denoted as Z_{th} combines the effects of the C_{th} thermal capacitance and R_{th} thermal resistance into a single metric. Briefly, the thermal impedance represents both the temporal and spatial changes of temperature in a heat conduction path, that is, the

[3]The work of Joseph Fourier, *The Analytical Theory of Heat*, published in 1822 inspired Georg Simon Ohm to formulate his law in 1827. For a long time, the heat spreading and the change of temperature, both known for many thousand years, helped to understand the newly introduced concepts of voltage and current through their analogy.

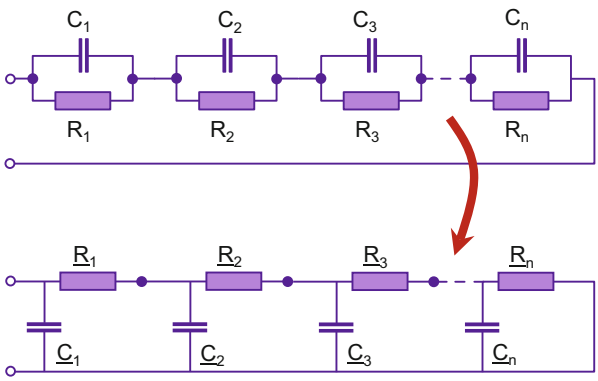

Fig. 2.5 Foster (**a**) and Cauer (**b**) type representation of discretized structures

thermal impedance is the ratio of the dynamically changing temperature and the dynamically changing power. There are multiple, equivalent representations of the thermal impedance.[4] These representations will be introduced step by step throughout the different subsections of this chapter. We start the discussion with the *network model representations* of the thermal impedance.

The theory makes a distinction between self-impedances and transfer impedances. Self-impedances express the response of the linear system at the location where the excitation occurs, simply saying at the *driving point*. Transfer impedances describe the system response at a different location often referred to as *monitoring* or *(thermal) test point*.

The network theory states that arbitrary complex *RC* network in which the excitation occurs between the single driving point and the ground behaves in the same way as a reduced set of thermal resistances and capacitances arranged in one of the configurations shown in Fig. 2.5. This corresponds to a thermal network model of a single powered junction and an isothermal ambient.

It is important to note that the internal nodes in these models generally cannot be associated with the monitoring points of a complex *RC* network.

The *chain of RC stages* style (top of Fig. 2.5) is called Foster representation, and the *ladder of RC stages* style (bottom of the figure) is called Cauer representation, named after the inventors who introduced these canonic circuit topologies into linear filter synthesis. The conversion between the two models is a standard procedure in network theory (summarized in Annex C of the JEDEC JESD 51-14 standard [40]). An example of the conversion is presented in GNU Octave (MATLAB) code below.

[4] In the theory of linear systems, impedances are interpreted in a stricter sense in the frequency (f, $\omega = 2\pi f$) domain which is connected to the time domain through the Fourier transform or in the complex frequency ($s = j\omega$) domain which is connected to the time domain through the Laplace transform.

Example 2.1: The Foster-Cauer Transformation

A simple MATLAB code which carries out the Foster-Cauer conversion is as follows:

```
% Foster to Cauer transformation
function [Rc Cc] = foster2cauer(Rf, Cf)
%calculate Z= p/q polynomial
n = length(Rf);
p = Rf(1);
q =  [Cf(1)*Rf(1) 1];
for k= 2:1: n
     pn= [0 p] + Rf(k)*q + [Cf(k)*Rf(k).*p 0];
     qn= [0 q] + [Cf(k)*Rf(k).*q 0];
     p = pn;
     q = qn;
end
% calculate Cauer form
for k= 1:1:n
     Cc(k) = q(1)/p(1);
     q= (q - Cc(k)*[p 0]);
     q(1)=[];
     Rc(k)= p(1)/q(1);
     p = (p - Rc(k)*q);
     p(1)=[];
end
end
```

Running the MATLAB code for the values in Fig. 2.21, one gets on the output:

```
[Rc Cc] = foster2cauer([1 1 4], [100e-6 10e-3 100e-3])
Rc = 1.0220 1.1870 3.7909;
Cc = 0.000098912 0.009159927 0.095994437
```

and

```
[Rc Cc] = foster2cauer([1 1 8], [100e-6 10e-3 100e-3])
Rc = 1.0220 1.1886 7.7894;
Cc = 0.000098912 0.009159713 0.093316320
```

2.3.2 Energy Balance and Stability

The simplest thermal model of a system consists of a single thermal resistance and a single thermal capacitance. This simplest model, shown in Fig. 2.6, consists of just two *thermal nodes*, the junction where power is applied and the ambient. The whole heat removal apparatus of the modeled system can be cumulated into a single and constant R_{thJA} junction to ambient thermal resistance. The energy storage is represented by a single C_{th} thermal capacitance.

The driving force of the heating and cooling processes is the thermal imbalance. Suppose there is a steady P_{gen} power generated in the system and fed into the "junction" node, which is at $T_J(t)$ junction temperature at t time. The temperature is always measured from a reference point; attributing any constant value to it in this simple model would not change the validity of the descriptive equations. In this section and in some further ones, we shall attribute $T_{amb} = 0$ temperature to the ambient in many cases, in order not to drag a constant value through all equations. This, of course, does not put any limitation to the validity of the equations.

We can state that before applying the P_{gen} power, the system is in a "low-power" thermal equilibrium, the energy stored in the system is zero, and so is the temperature of the T_J point. Applying P_{gen} power in the first instant elevates the internal thermal energy, $P_{gen} = P_{store}$ initially.

The continuous flow of P_{store} into the thermal capacitance increases the Q_{stored} thermal energy; thus, the T_J temperature elevates on C_{th}. We can recognize that P_{store} is the difference of P_{gen} and P_{diss}. Still, P_{gen} is constant, and the heat loss towards the ambient can be calculated as $P_{diss} = T_J/R_{thJA}$ at any time.

Solving the appropriate differential equations, we find that T_J grows exponentially:

$$T_J(t) = P_{gen} \cdot R_{thJA} \cdot (1 - e^{-\frac{t}{\tau}}), \tag{2.9}$$

where $\tau = R_{thJA} \cdot C_{th}$ is the characteristic *time constant* of the model.

Thus, during the transient process, the proportion of the heat loss through the junction to ambient thermal resistance grows, and the share towards the thermal capacitance diminishes.

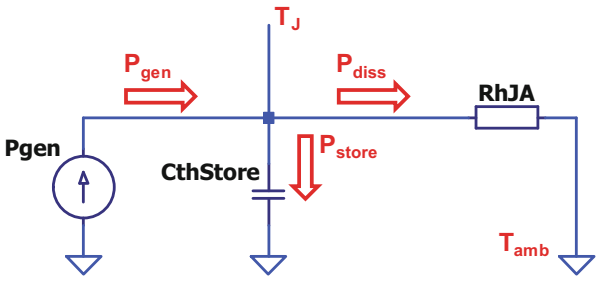

Fig. 2.6 Simple thermal model of a single RC stage driven by an external source. Generated P_{gen} and dissipated P_{diss} power indicated, their difference P_{store} elevates the Q_{stored} thermal energy

Integrating the heat flow (which is the power) over time the energy stored in and dissipated from the components can be also calculated. The integral of exponential functions until t time is straightforward, and the actual values can be directly derived from the electrical analogs.

The stored internal energy is $Q_{\text{stored}} = ½\ C_{\text{th}} \cdot T_{\text{J}}^2$, the applied energy is $Q_{\text{in}} = P_{\text{gen}} \cdot t$, and the dissipated energy is $Q_{\text{in}} - Q_{\text{stored}}$, applying the principle of conservation of energy. The principle corresponds to the Kirchhoff current law at the junction node in the interpretation of network theory.

Example 2.2: A Simple Thermal Model of a Device in Its Environment

When a thermal system is built for transient testing, it can be divided into a "device under test" (DUT) part and a thermal environment, a test bench. In this example the model of the DUT is simplified to two thermal resistances and two thermal capacitances, and the thermal environment is represented by a single thermal R_{th} and a C_{th} (Fig. 2.7).

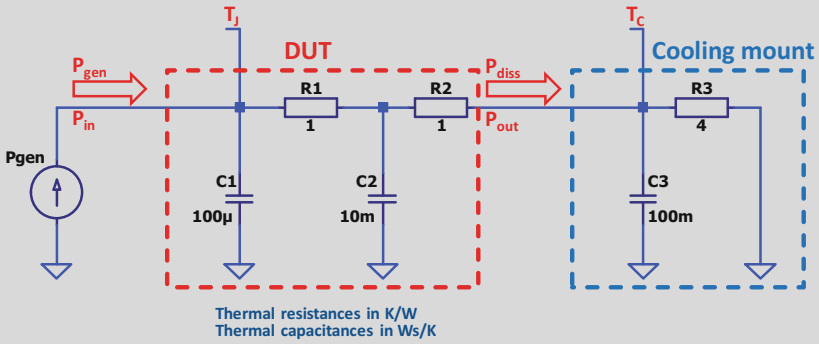

Fig. 2.7 Simple thermal model: a power device represented by two thermal RC stages, and its thermal environment, simplified to a single RC stage. Component values of network elements are assigned for further calculations

The temperature of the entry point (driving point) is denoted as T_{J}; the temperature of the separation point is denoted as T_{c}.

It should be emphasized that this highly simplified model resembles only superficially a realistic device with an exposed cooling surface, which is typically denoted as "case." The calculated temperature of such a separation point is related only loosely to an actual complex temperature distribution on the case of a physical device.

As the use of "dissipation" is often ambiguous, let us denote below the P_{gen} generated power by P_{in} and the P_{diss} dissipated power, the heat loss by P_{out}.

(continued)

Example 2.2 (continued)

In steady state the thermal capacitances do not influence the heat flow. The system can be characterized by a total R_{thJA} junction to ambient thermal resistance, $R_{\mathrm{thJA}} = R1 + R2 + R3$.

Steady state of the device is reached when $P_{\mathrm{in}} = P_{\mathrm{out}}$, and so the T_{Jss} steady state junction temperature is

$$T_{\mathrm{Jss}} = P_{\mathrm{in}} \cdot R_{\mathrm{thJA}}. \tag{2.10}$$

In the case when T_{J} was below T_{Jss} when P_{in} was applied, $P_{\mathrm{in}} > P_{\mathrm{out}}$ and heating occurs until steady state is reached. In the opposite case, $T_{\mathrm{J}} > T_{\mathrm{Jss}}$, then $P_{\mathrm{in}} < P_{\mathrm{out}}$ and a cooling process governs the system towards steady state.

In some cases, the heating process may end in a steady state where the T_{J} device temperature is above the absolute maximum ratings of the semiconductor. This thermal condition is called *thermal overload*.

Thermal overload results either in longtime degradation or in a sudden breakdown of the device; consequently it has to be avoided in normal device operation. This overload, however, can be intentional for reliability/lifetime testing purposes. By systematic reliability tests and analysis of the degradation mechanisms, the thermally influenced safe operating area (SOA) of devices can be established. A more detailed treatment of SOA definition is given in Chap. 6.

So far only such cases have been considered in which the input power is stable during the transient, and the heat removal can be characterized by a constant R_{thJA} junction to ambient thermal resistance in steady state. In Chap. 6 we examine the transients of devices which are normally operated at constant current during the test, and their electrical characteristics have positive thermal coefficient, PTC. This latter term means that their voltage grows with elevating temperature at constant current bias, resulting in an increase of their own, internal power generation.

In addition, also the temperature coefficient of the thermal conductivity can be positive in crucial portions of the thermal environment, in the heat-conducting path. This can result in growing junction to ambient thermal resistance at higher temperatures.

Either one of these effects or the coincidence of them can cause thermal instability; due to a positive feedback loop, the temperature keeps growing in the powered state, which also elevates the power. When this situation finishes only at extreme temperature causing fatal damage of the device, then we speak about *thermal runaway*.

The effects of thermal instability and thermal runaway are treated in Sect. 6.1 for diodes and in Sect. 6.2 for MOS devices.

2.3.3 Heating and Cooling Curves

With the help of the theory of linear systems, there is no need to restrict the waveform of an actual power change. For any of them, the corresponding temperature change and the relevant system descriptors can be derived easily. Further on in this book, however, we put special emphasis on those specific power profiles, which have an eminent role in thermal testing, namely, heating and cooling at constant applied power for prolonged time, or applied in a periodic manner.

The time domain response of a linear system to an arbitrary excitation can be calculated if some specific descriptive functions of the system are known. These are the $g(t)$ Green function, the response to a Dirac-δ excitation (approximated by a very short pulse of known energy), and the $a(t)$ response function to unit step excitation.

In case of an actual thermal transient measurement, in response to a $P(t)$ input power excitation of some waveform, the system will react with a $T(t)$ temperature response. For example, in the typical case when a constant power is applied to a previously unpowered system (step function like excitation), a monotonous temperature elevation can be observed until a new, "hot" steady state is reached.

Figure 2.8 shows the temperature elevation in a distributed thermal system, composed of a MOSFET device in a common package, mounted on a cold plate. This example will be used in further sections in this chapter to demonstrate how structural information of the assembly can be extracted from a simple thermal transient and how the information can be best visualized.

To a single short pulse of Δt length and P_p "height" in power, the system will react with a

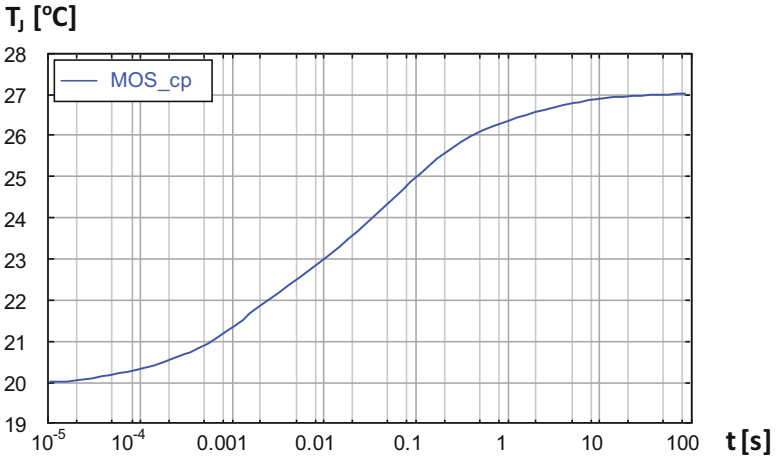

Fig. 2.8 Temperature change of a heated power device (MOSFET) on a water-cooled cold plate

$$T_{\mathrm{p}}(t) = \Delta t \cdot P_{\mathrm{p}} \cdot g(t) \tag{2.11}$$

response. Any continuous $P(t)$ power profile can be sliced into a sequence of short $\Delta t \cdot P_{\mathrm{p}}$ pulses. Linearity allows superposing the responses of the system, shifted in time as the pulses follow each other. Let us denote this series of pulses as $\Delta t \cdot P(t)$ where for the different t_i time instances $P_{\mathrm{p}_i} = P(t_i)$. With $\tau = (\Delta t \to 0)$, the series of the discrete P_{p_i} pulses will become a continuous $P(t)$ function. With this, according to the superposition principle, the $T(t)$ temperature change can be written as

$$T(t) = \int_{-\infty}^{\infty} g(t) \cdot P(t-\tau) d\tau \tag{2.12}$$

that is referred to as a *convolution integral*. Introducing the shorthand notation $\otimes$ for convolution, the above relation is written shortly as $T(t) = P(t) \otimes g(t)$. The formula in (2.12) expresses that at time t, the short pulse left behind by time τ is only present with weight $g(t)$ in the current temperature value.

There are strict conditions formulated for g and P in the theory of linear systems. These can be also expressed in ordinary terms. The $P(t)$ excitation has started at a certain time; the system was unpowered before that time. The retracting $P(t)$ function in (2.12) was zero before the start time; this can be reflected by changing the lower limit of integration in the formula from $-\infty$ to 0.

The convolution operation, $T(t) = P(t) \otimes g(t)$, also defines the components of the instrumentation needed for a transient measurement. A powering unit has to produce a $P(t)$ power profile during the measurement time. It is applied on the device under test, which is characterized by its $g(t)$ descriptive function. A data acquisition unit has to record the $T(t)$ temperature response. (As $g(t)$ embodies the relationship between the dynamic change of the temperature and the power, it can be considered as one possible theoretical representation of the thermal impedance.)

An obvious consequence of the above is that transient testing provides a direct means for system identification; the $P(t)$ excitation and the $T(t)$ response are known, in such a way the $g(t)$ system descriptor can be calculated:

$$g(t) = T(t) \otimes^{-1} P(t), \tag{2.13}$$

where $\otimes^{-1}$ denotes the deconvolution operation.

As in most cases no analytic solution exists for convolution and deconvolution, the results of these can be calculated by numerical methods only. Generic implementations of these operators are available in software libraries, MATLAB codes, etc. The numerical algorithms for deconvolution require much sophistication

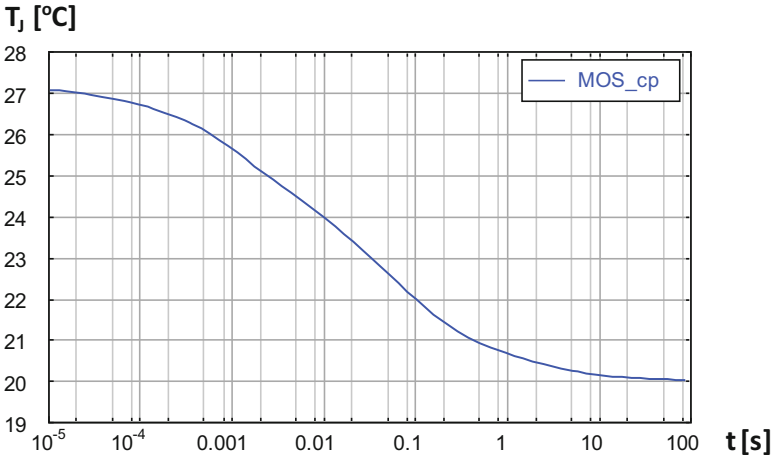

Fig. 2.9 Cooling of a power device (MOSFET) on a water-cooled cold plate, after heating is switched off in an abrupt, stepwise manner

and need careful considerations as advised in [58]. These considerations are typically overlooked in generic tools.

The linear approach implies that symmetric power profiles result in symmetric temperature changes. As a special case, heating and cooling processes are symmetric, if the previously mentioned nonlinearities do not apply.

The transient temperature change as illustrated in Fig. 2.8 or in Fig. 2.9 has a huge information content related to the structural details of the system, which is tested; but this information is not expressed in an obvious and apparent manner. The interpretation is much more evident in other equivalent representations highlighting one or other aspect of use. We can call these representations the "views" of a thermal transient.

In order to demonstrate the creation and use of the views, we shall present real transient measurements on real devices in actual thermal environments in the subsequent chapters. However, in some cases we found it more efficient to show the results in *simulation experiments* because it is easier to recognize the essence of a methodology in simplified structures. Moreover, these experiments can be easily repeated by the reader of this book.

Some simulated experiments will use the popular LTSpice analog simulator program [57] for analyzing both thermal and electronic properties of experimental arrangements. In these simulations the thermal equivalent of the electrical current is the heat flux ("flow quantity"), while the voltage carries the temperature values ("across quantity").

Example 2.3: A Simple Circuit Model of a Thermal System with Three Time Constants

A simple behavioral model of a thermal system is shown in Fig. 2.10, corresponding to the equivalent Foster network. In an electrical equivalent, the same current would flow through all stages in series; the voltage at the T_J driving point can be constructed as the sum of the voltage drops on the individual stages.

The thermal equivalent circuit in Fig. 2.10 has three time constants.

Running an LTSpice simulation with the circuit model, the "PULSE" directive forces $P_{in} = 10$ W input power at $t = 0$ time on the junction point. The simulation yields the T_J junction temperature transient response shown in Fig. 2.11.

We can observe three characteristic bumps in the curve; the reason of this will be discussed with more details in the next section.

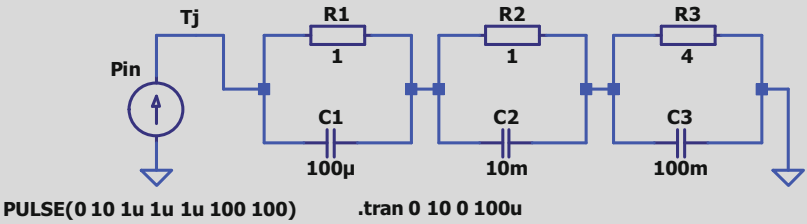

Fig. 2.10 Foster-type behavioral model of a simulated thermal system with three characteristic time constants ($\tau_1 = 100$ µs, $\tau_2 = 10$ ms, $\tau_3 = 400$ ms)

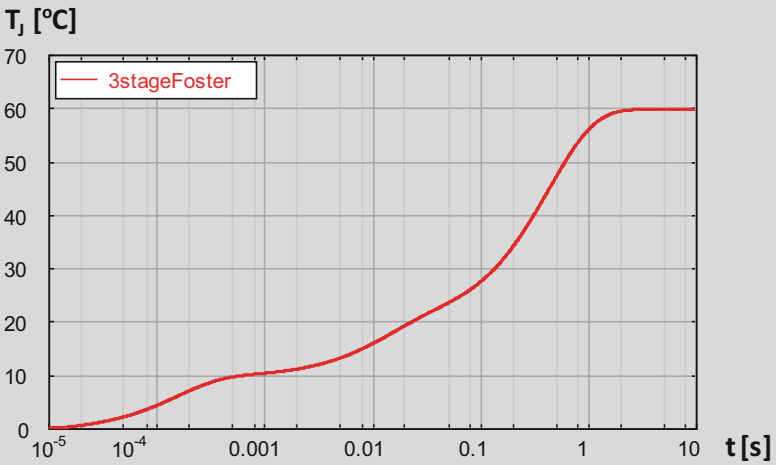

Fig. 2.11 Transient result from the LTSpice simulation defined in Fig. 2.10

Real thermal systems are sometimes simplified to similar circuits of three stages as a basic model. The first bump is interpreted as the heating of the chip with the die attach layer as obstacle in the heat spreading, the second bump corresponds to the heating of a package base with a thermal interface layer as bottleneck in the spreading, and the third one represents the characteristic heating of the cooling mount or the test environment. In the subsequent subsections, we shall present a more mature descriptive apparatus.

2.3.4 Z_{th} Curves

Examining Figs. 2.8 and 2.9 that present temperature transients measured on a real packaged MOSFET device, we can observe their "bumpy" nature.

These bumps are even more expressed in the simulated transient curve shown in Fig. 2.11 where the "system" has just three discrete time constants.

Engineering experience has proved that the structural information of the device under test and its thermal environment is encoded into the position and size of the bumps. In realistic measurements one can attribute the temperature change in the millisecond range to heat propagation in the die and through the die attach, in the second range to the cooling mount, in the minute range to heating of the circulated water, etc. This plot, however, characterizes the heat-conducting path only at the given powering.

The thermal conductivity and specific heat of the device components and of the measurement environment show only minor change in the typical temperature range of use. This implies that shifting the base plate temperature, we obtain similar recorded curves, and altering the applied power, we obtain again similar, proportionally magnified records.

Normalizing the temperature change with the applied power, we obtain the Z_{th} *curves* (Fig. 2.12). Sometimes the Z_{th} curve is referred to as the *thermal impedance curve.*[5]

At this point it has to be noted that it is common in the engineering practice that quantities which change over many orders of magnitude like time or thermal capacitance are plotted along logarithmic axes. Still the axes are labeled with the values of the quantity in the original, linear scale. In accordance with the customary representation, we refer to these quantities as $T(t)$, $C_{\text{th}}(R_{\text{th}})$, although the plot corresponds to $T(\log t)$, $\log C_{\text{th}}(R_{\text{th}})$, and so on. This also applies to the Z_{th} chart in Fig. 2.12 and all subsequent similar ones.

Z_{th} curves are always *monotonically increasing* due to their definition, because a heating curve (Fig. 2.8) is normalized with a positive and a cooling curve (Fig. 2.9) is

[5] In electronics the impedance is primarily defined in the frequency domain, not in the time domain as a step response function.

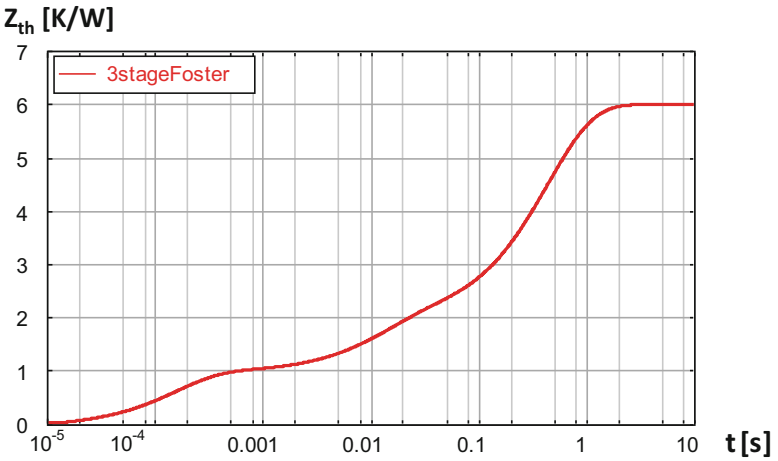

Fig. 2.12 Temperature change normalized by the applied power: the Z_{th} curve of the thermal system introduced in Fig. 2.10

normalized with its negative power step. The result is nearly the same Z_{th} curve in the two cases.

As a fairly accurate temperature transient for any power step can be produced if we multiply the Z_{th} curves by the actual power applied, this curve is used frequently for the characterization of the thermal behavior.

This concept of proportionality to power (i.e., linearity) is not fully accurate when measuring realistic systems. The actual shape of a cooling or heating curve depends on the temperature dependence of the material parameters as well. A more dominant factor is that at increased power level and at higher temperature elevation, the cooling mechanisms (turbulent convection, radiation) become more intensive; consequently the real temperature change is lower than the one extrapolated from the multiplied Z_{th} curve. As such, using Z_{th} for temperature estimation, let us remain on the safe side.

A deeper analysis of nonlinear effects is given in [64].

2.4 System Properties Calculated from the Thermal Transient

2.4.1 Time Constant Spectra

In a both theoretically and practically important case, a constant P_{on} power is turned on and kept on the tested system for a prolonged time.

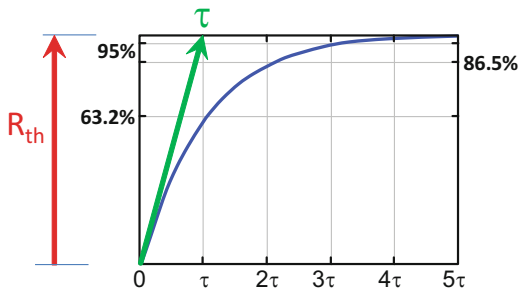

Fig. 2.13 Time response of a single thermal RC stage to a step function excitation with its magnitude and time constant shown

For unit step powering which is zero before switching on and unit value afterwards, based on the theory of linear systems, (2.12) will be of the form of

$$a(t) = \int_0^t g(x) \cdot 1 dx. \tag{2.14}$$

This means that the $a(t)$ unit step response is the integral of the $g(t)$ Green function. Producing a power step in reality is much easier than generating extremely short high energy pulses. Instead of directly measuring it, the $g(t)$ function can be derived from the measured step response. Eq. (2.14) implies that

$$g(t) = \frac{d\, a(t)}{dt}. \tag{2.15}$$

The unit step response function, for which the traditional notation in linear system theory is $a(t)$, is exactly the function that is called the $Z_{\text{th}}(t)$ normalized temperature transient curve in thermal engineering practice, presented earlier. In this book we use both notations.

When a constant P_{on} power is switched on at zero time, a single RC stage in Fig. 2.10 produces an exponential growth after switching on the power, adding a $T(t) = P_{\text{on}} \cdot R_{\text{th}} \cdot (1 - e^{-t/\tau})$ temperature term to the response of the entire series of the RC stages (Fig. 2.13). In the analogous electronic circuit, a steplike current is switched on, and exponential voltage growth is observed at the node driven by the current source.

The R_{th} *magnitude* denotes the thermal resistance of an elementary stage; $\tau = R_{\text{th}} \cdot C_{\text{th}}$ is the *time constant* where C_{th} is its thermal capacitance. Adding up the temperature drops of each subsequent stage in the series model, at the input (in this case at the junction), we get a *sum* of exponential functions:

$$T(t) = P_{\text{on}} \cdot \sum_{i=1}^{n} R_{\text{th}_i} \cdot (1 - e^{-t/\tau_i}). \tag{2.16}$$

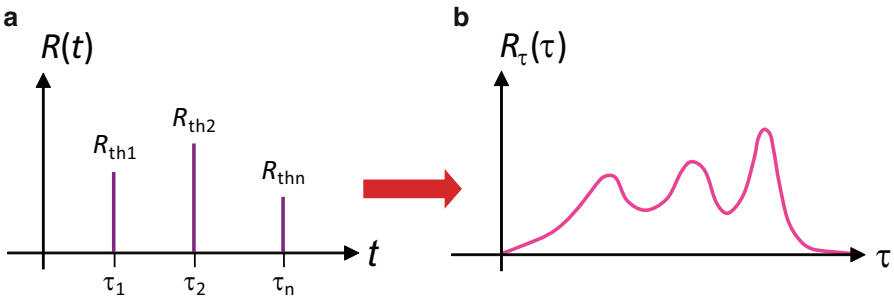

Fig. 2.14 Time constants in a lumped element system (**a**) and in a distributed parameter system (**b**)

Normalizing the $T(t)$ temperature response with the P_{on} power, we obtain the $Z_{\text{th}}(t)$ *thermal impedance*:

$$Z_{\text{th}}(t) = a(t) = \frac{T(t)}{P_{\text{on}}} = \sum_{i=1}^{n} R_{\text{th}_i} \cdot (1 - e^{-t/\tau_i}) \tag{2.17}$$

that is nothing else than the $a(t)$ unit step response function, introduced earlier.

Different material slices have different characteristic thermal parameters; the resulting different R_{th_i} magnitudes and τ_i time constants make the $T(t)$ curve "bumpy," as seen in Figs. 2.8 and 2.11. The bumps are originated from the individual time constants (Fig. 2.14a) of the system.

The system can be fully characterized by a proper number of τ_i and R_{th_i} pairs; the sum in (2.16) restores the exact waveform of the temperature at the driving point. As the equation indicates, the dimension of magnitudes is K/W (kelvin/watt), and the dimension of the time constants is second.

If the thermal system is subdivided into a large number of thin slices, we reach a *continuous* model of elementary sections, forwarding energy into each other as indicated in (2.2) and (2.3). The sum formulae in (2.16) and (2.17) take the form of an integral:

$$T(t) = P_{\text{on}} \cdot \int_0^{\infty} R_\tau(\tau) \cdot (1 - e^{-\frac{t}{\tau}}) d\tau. \tag{2.18}$$

$$Z_{\text{th}}(t) = a(t) = \frac{T(t)}{P_{\text{on}}} = \int_0^{\infty} R_\tau(\tau) \cdot (1 - e^{-\frac{t}{\tau}}) d\tau. \tag{2.19}$$

We can easily realize that these relationships resemble the form of the convolution integral shown by Eq. (2.12).

Equation (2.18) corresponds to a measurement scheme again, in a similar but not equivalent way as it was found in (2.12) before. After the test equipment applies a P_0 constant power on the system under test at zero time, a measured $T(t)$ temperature response is recorded. The $R_\tau(\tau)$ time constant spectrum is now the characteristic system descriptor. As stated above for the discrete τ_i and R_{th_i} pairs, also the integral in (2.18) can be used for restoring the temperature waveform.

Although the Eqs. (2.18) and (2.19) seem to operate on continuous functions, the numerical procedures to obtain them produce discretized results. The primary parameter for the procedure is the intended $\Delta\tau$ time constant step.

Selecting small $\Delta\tau$ results in a high number of RC pairs in the discretized $R_\tau(\tau)$ time constant spectrum. The corresponding Foster-type network models properly the dynamic thermal behavior of the measured heat-flow path, seen from the T_J driving point (Fig. 2.15). The R_{th_i} magnitude element in the chain is composed as the product of the $R_\tau(\tau_i)$ spectrum value and the width of the τ_i time constant slice.

The C_{th_i} thermal capacitances can be calculated from the $\tau_i = R_{\text{th}_i} \cdot C_{\text{th}_i}$ relationship.

Real physical objects, which are to be tested as thermal systems, obey mechanical constraints. The heat propagates from a very thin active layer through tiny semiconductor chips, which are mounted into larger packages and modules. These devices are mounted on larger heat sinks, located in enclosures. Accordingly, thermal time constants of an actual electronic system may range from microseconds (thermal transient within the chip) to hours (temperature elevation of a cooling mount). For this reason are transient curves plotted in this chapter in logarithmic time scale; otherwise, the tiny details of early times could have been lost.

Similarly, thermal transient testers record the transients in a logarithmic time scale in order to store the transient results in a manageable amount of data. It is reasonable

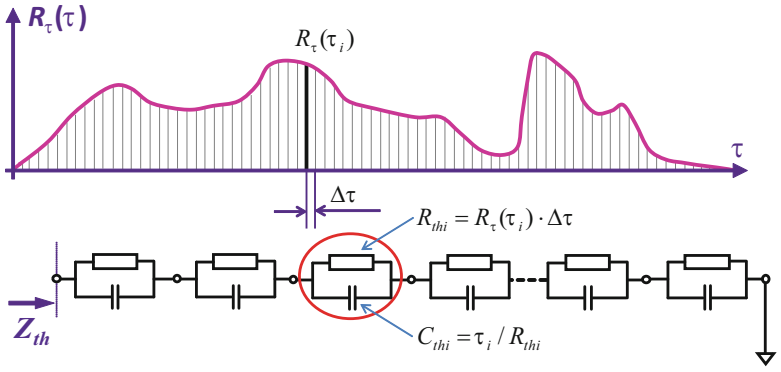

Fig. 2.15 Discretized time constant spectrum with equidistant $\Delta\tau$ time constant steps and the equivalent Foster chain

to process these data sets in *logarithmically equidistant* time increments, which correspond to a constant ratio of increments in linear time.

Introducing the $z = \ln(t)$ transformation for the time and the $\zeta = \ln(\tau)$ transformation for the time constants has a further advantage; it offers an easy transition between the convolution-style system descriptor in (2.12) and (2.13) and the "sum of exponentials" style system descriptor in (2.19).

The R_ζ time constant spectrum[6] defined on a logarithmic time scale can be obtained, indeed, in a convolution-type relationship with the first derivative of the unit step response. That is, the

$$\frac{d}{dz}Z_{\text{th}}(z = \ln(t)) = \frac{d}{dz}a(z) = R_\zeta(z) \otimes w_z(z) \tag{2.20}$$

relationship holds where w_z is a fixed function: $w_z(z) = \exp[z - \exp(z)]$. (See further details in [58, 59].)

Equation (2.20) tells that from the unit step response of a (thermal) system, its time constant spectrum can be extracted by deconvolution as follows:

$$R_\zeta(z) = \left[\frac{d}{dz}a(z)\right] \otimes^{-1} w_z(z). \tag{2.21}$$

At this point we have to emphasize that formulating (2.19) and transforming it in some steps to (2.21) is not aimless equation crunching; it is the essence of testing and system identification. Eq. (2.19) corresponds to the scheme of thermal testing; the thermal tester applies P_{on} on the tested system and records the $T(t)$ temperature response. Eq. (2.21) depicts a systematic process, which yields an R_ζ system descriptor, fully characterizing the thermal behavior of the system.

R_ζ is calculated from the $a(z = \ln(t))$ system response and a *fixed* auxiliary w_z function.

The deep mathematical background of the calculus and the fundamental concepts related to the time constant spectra of distributed RC systems are provided in [58–60].

The systematic use of the deconvolution which starts from the measured transient and results in the time constant spectrum calculated by (2.21) is the broadly used *network identification by deconvolution*, shortly the *NID method*.

The disadvantage of the Foster type of network model is that, although it is a *mathematically correct* model of the transient behavior, it cannot be used for building an equivalent of the real, *physical* thermal structure, because it contains *node-to-node* capacitances.

In the fundamental heat transfer equation (Eq. (2.7)), the T_1 and T_2 temperatures are measured from the ambient; a C_{th} thermal capacitance exists *between* a point representing the material portion and the *ambient* as underscored by (2.1).

[6] In contrast to the previous R_τ notation, with the index in the R_ζ function, we indicate that this is the version of the time constant spectrum that represents the time constant distribution on the logarithmic time scale.

Accordingly, the real thermal capacitances are always connected to the ground, since the stored thermal energy, which they represent, is proportional to the temperature elevation of *one node* with respect to the reference, and not to the temperature difference of two nodes as would be suggested by the Foster model. For this reason, the calculated Foster model has to be transformed into a Cauer model with a standard mathematical transformation, as outlined in [58, 61], and Annex C in the standard [40].

In order to facilitate understanding, let us examine the above-discussed functions of an artificial system.

Example 2.4: The Characteristic Functions of a Known System

A simple thermal system and its step response were presented in Fig. 2.10 and in Fig. 2.12. As the system is an artificial one with *exactly* three time constants, we can also produce the signal constituents with an appropriate simulation. The circuit and its separated subcircuits are shown in Fig. 2.16.

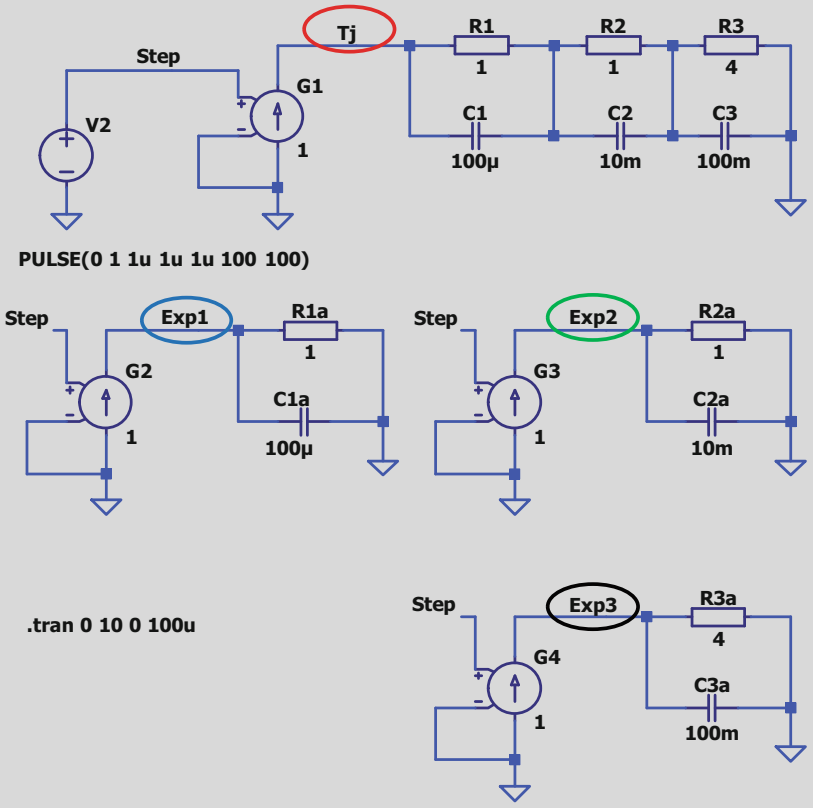

Fig. 2.16 Equivalent circuitry in LTSpice for producing the T_J temperature of the simulated system and for directly providing the three constituents of the resulting curve

(continued)

Example 2.4 (continued)

In Fig. 2.17 we present the simulated Z_{th} curve of Fig. 2.12 again, with the three bumps corresponding to the time constants and magnitudes, and also the constituting three exponential curves as *Exp1*, (R1 = 1 K/W, $\tau 1$ = 100 µs); *Exp2*, (R2 = 1 K/W, $\tau 2$ = 10 ms); and *Exp3*: (R3 = 4 K/W $\tau 3$ = 400 ms).

It can be observed that the Foster representation gives some information on the nature of the Z_{th} curve, the time constants correspond to some extent to the position of the bump, while the magnitude refers to the curvature at that location.

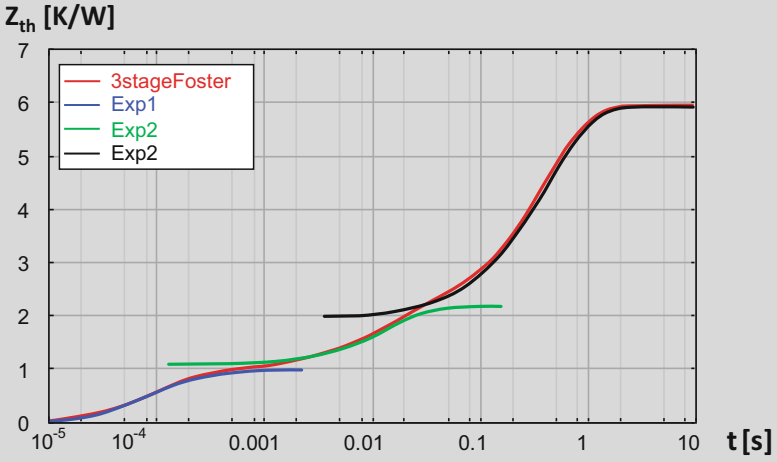

Fig. 2.17 Z_{th} curve with the three exponential components shown

If we calculate back the time constants from the simulated thermal transient curve with the NID method; we obtain the time constant spectrum shown in Fig. 2.18.

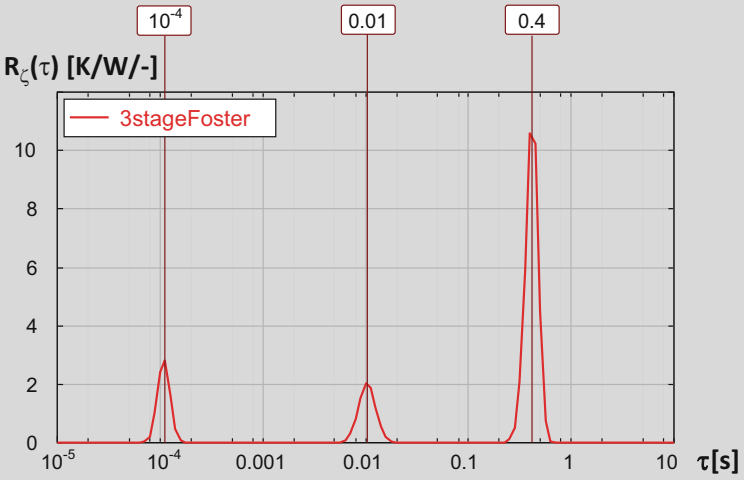

Fig. 2.18 Time constant spectra acquired by the NID method in the three-element lumped element system

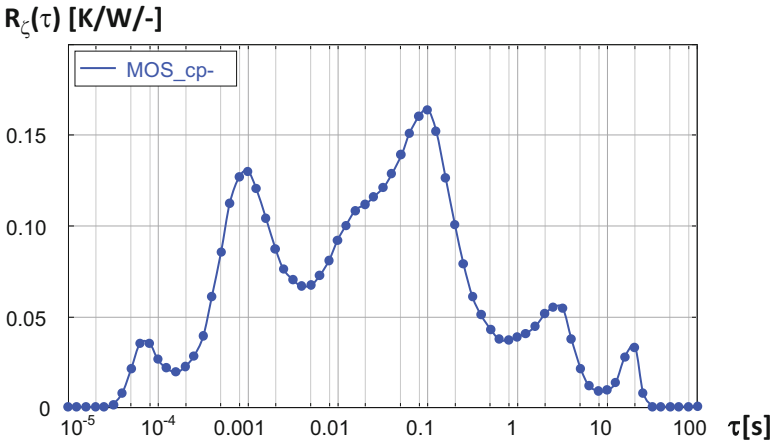

Fig. 2.19 Time constant spectrum of a real distributed parameter system (MOSFET on cold plate) calculated from the measured transient of Fig. 2.8

Real systems have many time constants forming a quasi-continuous time constant spectrum. When discrete time constants of an artificial system are calculated in an *iterative realization* of the deconvolution, they appear as peaks in the time constant spectrum, smashed a bit around a center point, due to different inherent limitations of any numerical process. The resolution of the results obtained by the NID method is also limited by the noise inherently present in the signals. Even in simulated curves (that can be extremely accurate), there is always a quantization noise present. Note however that despite limitations of the resolution, the locations of the maxima of the time constant spectrum such as the ones shown in Fig. 2.18 are exactly at the same time constant values as calculated directly from the element values of the Foster-type network model presented in Fig. 2.16. A deeper insight into the fundamental limitations in the resolution of the restored time constant spectra is provided in [58].

Applying the NID method on the transient measurement result of Fig. 2.8, the time constant spectrum of Fig. 2.19 was obtained.

The real distributed system (MOSFET on cold plate) consists of many elementary portions of matter forwarding and storing energy. The result of discretization can be observed in the plot; the time scale was split into 75 equidistant $d\zeta$ slices in logarithmic time. Further reading on generating time constant spectra and verifying their correctness is available in [174, 175].

2.4.2 The Structure Functions

Discrete time constants or a time constant spectrum can be produced from a measured $T(t)$ curve with the process presented so far. Still, the equivalent Foster circuit is a behavioral model[7] only.

In reality however, a slight change at a material layer in an actual physical object distorts many time constants in the chain. A physically sound approach can be based on the equivalent Cauer model of the system where we have a chain of RC stages with nodal capacitances, as illustrated in Fig. 2.20. In many cases elements of such a

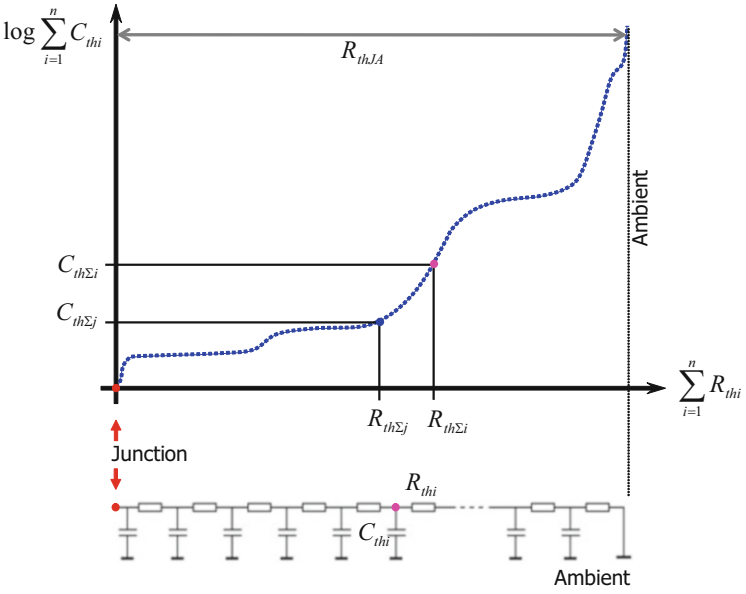

Fig. 2.20 Structure function: the graphic representation of the thermal RC equivalent of the system

[7]The behavioral nature of the Foster model can be endorsed by several arguments. As defined in Sect. 3.1, a thermal capacitance component in the network model of a thermal system represents the heat storage capability of a material region, lumped into a node of the network model. It describes the temporal changes of the temperature of that node with respect to the ambient; hence, it is referred as *nodal capacitance*. Accordingly, in an RC network model of a thermal system, it is represented by a thermal capacitance between a node of the model and the thermal reference point (ambient). Since the Foster-type models consist of *parallel* RC stages between adjacent nodes, neither the thermal resistances nor the thermal capacitances in it can be associated directly to the physical structure of the heat-flow path.

Equation (2.16) represents the Foster model in a mathematical formula. One can sum up the constituents in any order; the way in which sequence the individual stages are placed into the chain does not affect the overall temperature response at the junction. Based on the simple principle of mechanics that the junction is the tiniest "part" in the heat-conducting path and then smaller constituents are always mounted on larger ones, during the Foster-Cauer conversion, the stages are processed in growing order of $\tau_i = R_{\mathrm{th}_i} \cdot C_{\mathrm{th}_i}$ thermal time constants.

model can be attributed the different portions of the heat-flow path structure; changing the order of two RC stages in a Cauer model results in a different $T_J(t)$ junction temperature response. That is, a Cauer-type network model is not only another behavioral model of a thermal system, but it is also characteristic to the thermal system like a *signature*. Based on the so-called frequency domain calculus of the linear RC networks, there are standard procedures for the Foster ↔ Cauer model transformations (see, e.g., Annex C of [40]).

Thus, a direct synthesis method for transforming the measured junction temperature transients into a *compact model* of the physical structure of a complex 3D thermal system is provided by accomplishing the deconvolution process, then discretizing the obtained thermal time constant spectrum and converting it into a Foster-type network model, and completing the Foster → Cauer transformation in a sequence. An early formulation of the concept was given in [137] for analogous mechanical problems, a modified method is presented in [136].

The obtained compact thermal model is a direct, physical representation of heat-flow path sections in which the heat spreading occurs in a true *one-dimensional* (1D) manner. Moreover, in cases where the spreading pattern can be expressed as a simple function of a single space coordinate, introduced as *essentially 1D* spreading in Sect. 2.5, the physical structure can be identified in a similar way. In all cases, regardless of the 1D, essentially 1D or complex 3D nature of the actual heat-spreading pattern, the Cauer-type models and their further representations are the *unique thermal signatures* of the physical structure of a semiconductor device package.

The first equivalent representation of a Cauer-type RC ladder model describing the heat-flow path is a graph, called *structure function* (shown in Fig. 2.20).

The quantities shown on the axes in the figure are the *cumulative thermal resistance*, defined as

$$R_{\text{th}\Sigma} = \sum_i R_{\text{th}i} \tag{2.22}$$

and the *cumulative thermal capacitance*

$$C_{\text{th}\Sigma} = \sum_i C_{\text{th}i} \tag{2.23}$$

In other words, starting from the driving point (the junction), we cumulate (sum) the partial thermal resistance and thermal capacitance values for of all subsequent heat-flow path sections. If we interpret the cumulative thermal capacitance as function of the cumulative thermal resistance, we obtain the *structure function* sometimes called *cumulative structure function*, often abbreviated as *SF*:

$$SF = C_{\text{th}\Sigma}(R_{\text{th}\Sigma}) \tag{2.24}$$

The structure function is an excellent graphical tool to visualize the heat-conducting path. In accordance with the ladder of the figure, this chart sums up the thermal resistances, starting from the heat source (junction) along the x-axis and the thermal capacitances along the y-axis.

In *low gradient sections* of the structure function, a small volume, representing small thermal capacitance, causes large change in the thermal resistance. These regions have either *low thermal conductivity* or *small cross-sectional area*, or both. *Steep sections* in the function correspond to material regions of either *high thermal conductivity* or *large cross-sectional area*, or both. Sudden breaks of the slope belong to material or geometry changes.

Thus, thermal resistance and capacitance values, geometrical dimensions, heat transfer coefficients, and material parameters can be directly read on structure functions.

It is sometimes easier to identify the interface between the sections using the derivative of the cumulative curve: the *differential structure function*. Here peaks correspond to regions of high thermal conductivity like the chip or a heat sink, and valleys show regions of low thermal conductivity like die attach or air. Interface surfaces are represented as inflexion points between peaks and valleys.

From (2.6) and (2.8), we can say:

$$DSF = \frac{dC_{\text{th}\Sigma}}{dR_{\text{th}\Sigma}} = c_{\text{V}} \cdot dx \cdot A \cdot \left(\frac{1}{\lambda}\frac{dx}{A}\right)^{-1} = c_{\text{V}} \cdot \lambda \cdot A^2 \tag{2.25}$$

The *differential structure function* (frequently abbreviated as *DSF*) yields information on the cross-sectional area along the heat conduction path. Further reading on producing structure functions and verifying their correctness is available in [176, 177].

In order to show how structural changes are represented in a structure function, we analyze below a simple artificial thermal model used in Example 2.3.

Example 2.5: The Structure Function of a Model Network

In order to demonstrate the easy usability of the structure functions, let us consider the following, still artificial example. In Fig. 2.21 we present two Cauer-type RC networks. The upper one is the converted Cauer-style version of Fig. 2.10. It can be observed that the thermal resistance and capacitance values slightly differ from the ones in the original Foster network, but based on the linear network theory, they produce equivalent thermal response. In other words, they have identical driving point impedance.

For making the example more plausible, we divided the three stages into two portions again: we assigned two stages to the device model; the third represents the test bench.

(continued)

Example 2.5 (continued)

Identification of structural elements in a system can be best facilitated with intentional material or geometry changes at relevant structural interfaces.

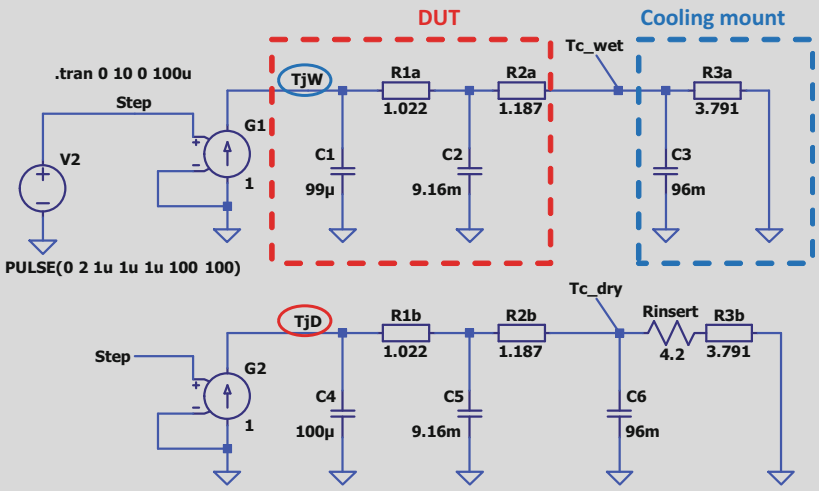

Fig. 2.21 The circuit scheme of Fig. 2.10 converted to Cauer equivalent. Two stages represent the device model; the third is the model of the test bench. Different thermal interface material quality is considered in the difference of the thermal resistance towards the ambient

To demonstrate this, the circuit scheme was altered to express the effect of different thermal interface materials between the device under test (DUT) and the test bench or cooling mount part. The different material quality was modeled by changing the 3.791 K/W thermal resistance towards the ambient to approximately 8 K/W in the second circuit.

The model of Example 2.5 resembles the concept of the broadly used JEDEC JESD 51-14 thermal measurement standard [40], also called the transient dual interface method (TDIM), developed for the testing of devices with an exposed cooling surface, such as packages with a cooling tab or modules with a baseplate. This separates the device under test from the cooling environment based on two thermal transient measurements, once on a cold plate without applied thermal paste ("dry" boundary) and then with appropriate wetting with a high-quality thermal paste material ("wet" boundary). This latter scheme corresponds to the actual intended mounting of power devices.

The inserted 4.2 K/W thermal resistance for the "dry" case can be interpreted as the thin air gap between the device and the cooling mount on a dry surface.

In the simulation example, 2 W power step was applied at $t = 0$ to the **TjW** and **TjD** input points. The simulated thermal transient curves are shown in Fig. 2.22.

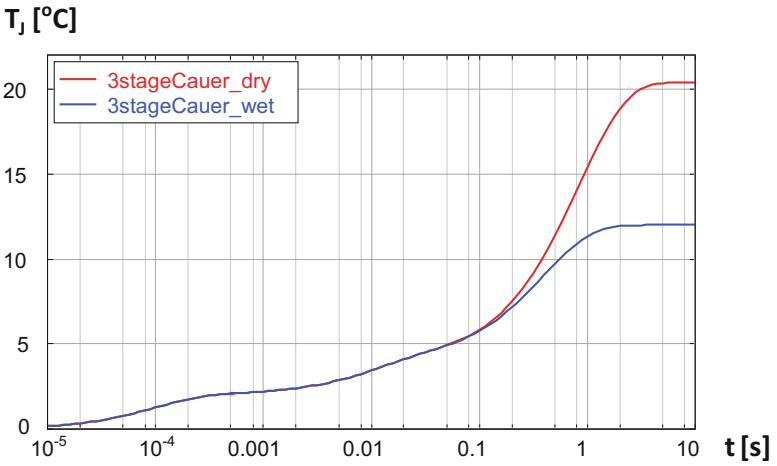

Fig. 2.22 Simulated temperature response of the two system variants at 2 W applied power

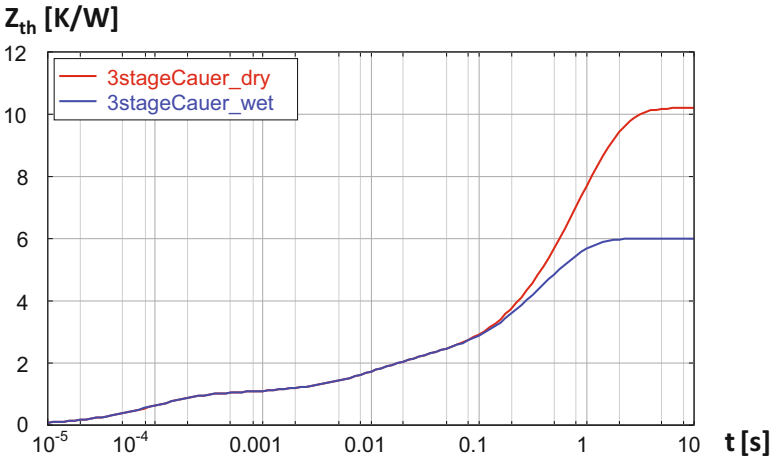

Fig. 2.23 Z_{th} curves of the two system variants

We can see in the figure that the two curves overlap up to about 0.2 s, but after that the junction temperature in the "dry thermal interface" case starts to increase much faster than in the "wet" case.

The same phenomenon can be observed of course also on the Z_{th} curves.

The time constant spectra of Fig. 2.24 are generated from the Z_{th} curves of Fig. 2.23 using the T3Ster-Master standard transient evaluation tool of a thermal tester equipment [54].

The simulated network represents a discretized system with three discrete time constants, while real systems are always continuous ones. The thermal transient evaluation software carries out mainly the steps defined in the previous sections.

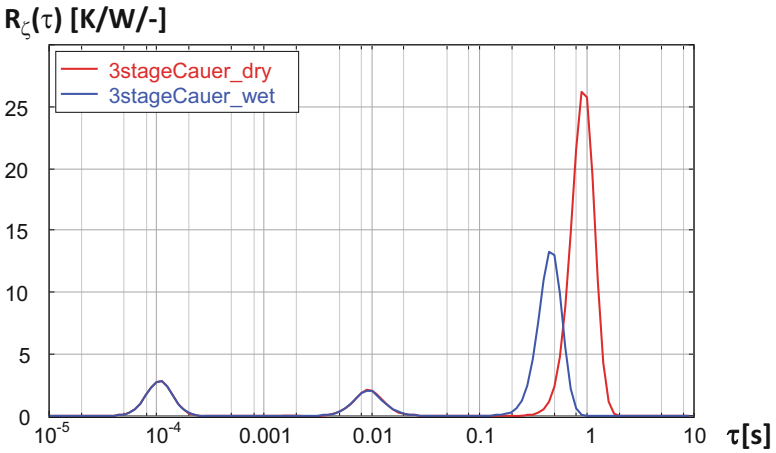

Fig. 2.24 Time constant spectra of the two system variants

Accordingly, it produces a continuous spectrum as defined in (2.21). A discretized time constant set can be produced from it by identifying the position of the three peaks and interpreting the area under the peaks as its magnitude.

There are various ways to accomplish the deconvolution process. A frequently used methodology is based on an iterative algorithm [58]. The actual deconvolution process resulted in the curve of Fig. 2.24 in 1000 iterative steps. Higher iteration number would result in sharper peaks around the time constants but with the same magnitude.

In Fig. 2.25 the structure functions of the thermal network of Example 2.5 are shown. The steep elevations correspond to the resistive elements and the flat plateaus to the capacitances in the Cauer-type representation. The readout of the values was performed with manual cursor positioning in the evaluation software [54] on the elevations and plateaus; an insignificant difference of these measured values from the original values can be observed.

It can be observed that the curves belonging to the "wet" and the "dry" boundary condition mostly coincide until the heat propagation in the "DUT" part in our example of Fig. 2.21 is represented, showing the fact that the two cases differ only in the last part of the heat-flow path.

The partial thermal resistances around 1 K/W each and the total R_{thJA} junction to ambient thermal resistance, 6 K/W for the "wet" and 10 K/W for the "dry" boundary, can be easily identified, and so are the appropriate thermal capacitances.

In real structures the steps in the structure functions are obviously less expressed, as demonstrated later in Example 2.7.

In Fig. 2.26 the differential structure functions of the thermal network and the manually measured peak positions, corresponding to the highest steepness in the cumulative structure function, are shown.

The next examples help in understanding the use of the structure functions.

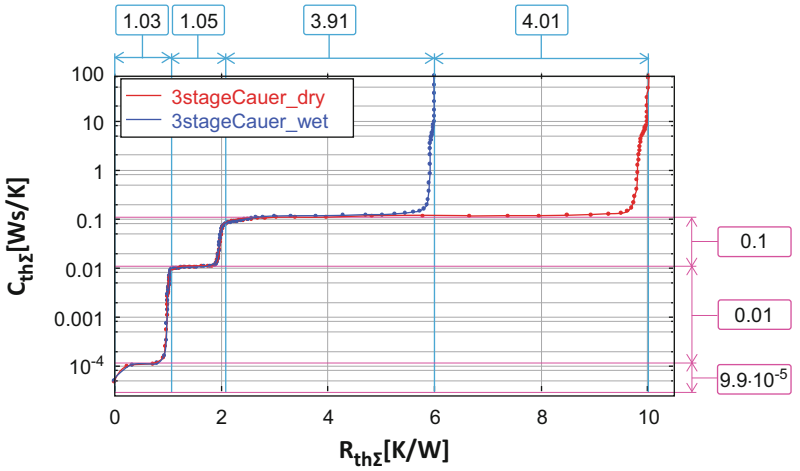

Fig. 2.25 (Cumulative) structure functions of the thermal network in Fig. 2.21. The steep elevations correspond to the resistive elements and the flat plateaus to the capacitances in the Cauer-type representation

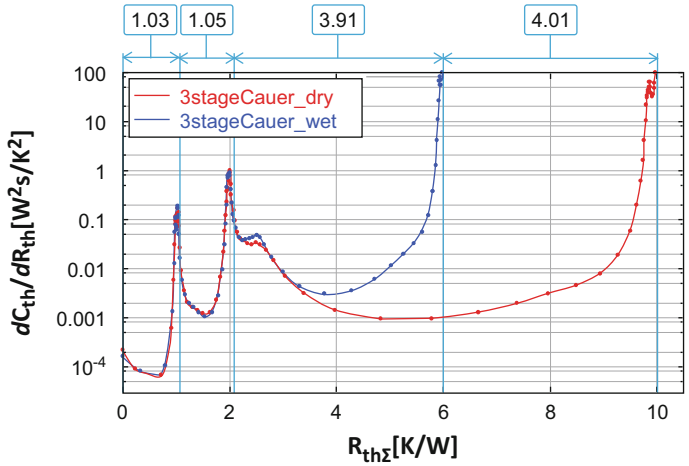

Fig. 2.26 Differential structure function of the thermal network in Fig. 2.21. The peak positions correspond to the resistive elements in the Cauer-type representation

Example 2.6: Analysis of the Heat Transfer in a Homogeneous Rod

A homogeneous rod with thermal boundary conditions is shown in Fig. 2.27. This rod can be considered as a series of infinitesimally small material sections as discussed above. Consequently, its discretized network model would also

(continued)

Example 2.6 (continued)

be a series connection of the single RC stages as shown in the figure. Thus, with this slicing along the heat conduction path, we create a ladder of lateral thermal resistances between two thermal nodes, and thermal capacitances between a node and the ambient.

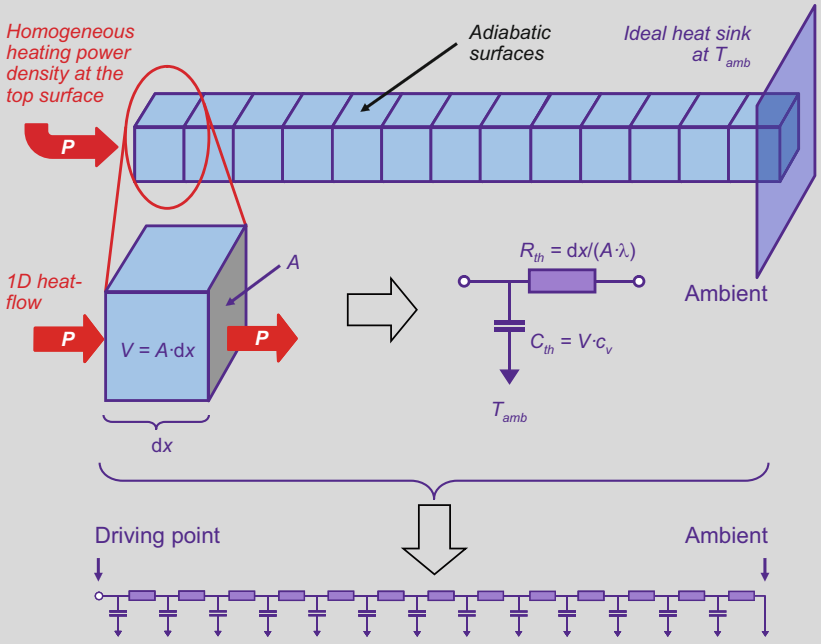

Fig. 2.27 The RC model of a narrow slice of the heat conduction path with perfect one-dimensional heat flow and the Cauer-type network model of the thermal impedance of the entire heat-flow path

Since homogeneity is assumed, the ratio of the elementary thermal capacitances and thermal resistances in the network model shown in Fig. 2.27 is constant. This means that the structure function of the rod is a straight line – its slope is determined by the C_{th}/R_{th} ratio of the network model and its differential structure function would be a constant – equal to the C_{th}/R_{th} ratio of the element values, as shown in Fig. 2.28.

This rod example demonstrates that the features of the structure functions are in a one-to-one correspondence with the properties of the heat conduction path.

Let us assume that in a given section in the middle of the rod, the C_{th}/R_{th} ratio is doubled. This results in a steeper middle section in the cumulative structure function (with the slope doubled) and in a peak in the differential structure function (which is twice as high as the constant value of the other sections). This is illustrated in Fig. 2.29.

(continued)

Example 2.6 (continued)

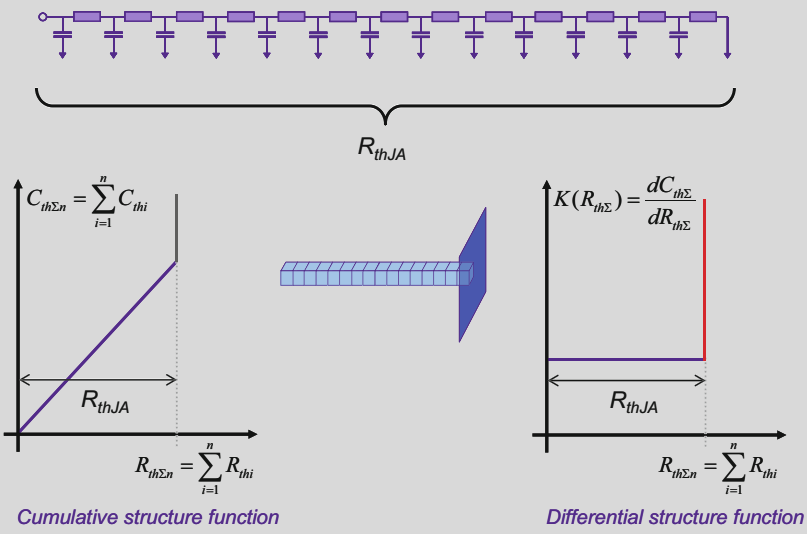

Fig. 2.28 The cumulative and differential structure functions of a homogeneous rod

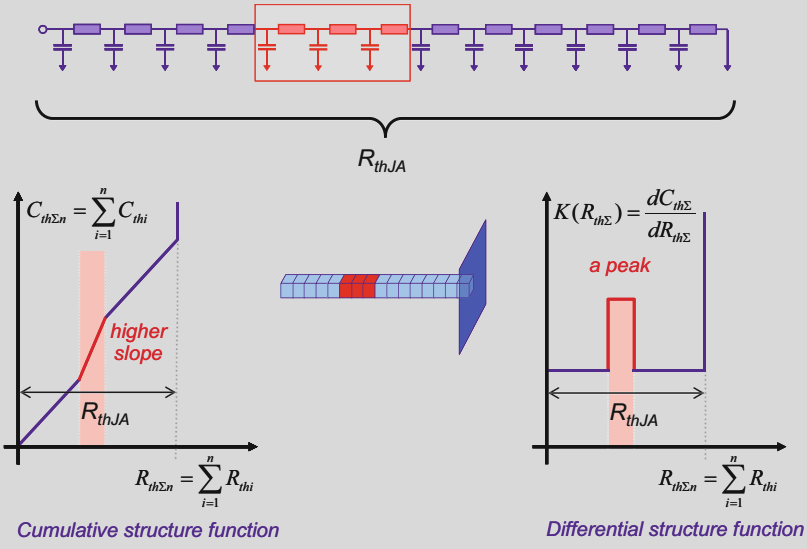

Fig. 2.29 The structure functions indicate the changes in the C_{th}/R_{th} ratio along the heat conduction path

(continued)

Example 2.6 (continued)

Obviously, this "doubling" of the C_{th}/R_{th} ratio can be the result of reaching a material section with different λ or c_V material parameters, or of larger cross-sectional area.

This discretized model offers a way to determine the thermal behavior of the rod. At any driving point excitation, an analog simulator tool, e.g., [57] solving the response of the RC ladder yields the temperature of any point within the rod, as it varies in time.

It is important to notice that there is no one-to-one correspondence between the number of material layers in a laminate structure and the number of time constants assigned to the system. Strictly taken, a homogeneous rod also has an infinite number of time constants, which can be calculated with (2.21). In the frequent case, however, when a material stack is composed of bulky layers of high conductance and thin layers of low conductance, a few characteristic time constants can be identified from the capacitance of the bulky layers and the resistance of the thin layers. A Cauer → Foster backwards transformation yields the major time constants from the identified thermal resistances and thermal capacitances.

Still, when the smallest thermal time constant of a system is to be ascertained, a generally good estimation can be the thermal capacitance of the chip multiplied by the thermal resistance of the die attach layer.

The RC chain normally starts with a capacitance, in order to avoid a temperature elevation of infinite steepness as a response to a sharp power step. An alternative composition of the Cauer ladder is proposed in [9]; at higher number of constituents, that approach is equivalent to the scheme outlined in Fig. 2.29.

Due to its simplicity, the heat spreading problem in the homogeneous rod has known solutions also in the continuous approach. The heat equation which was presented in various forms from (2.3) to (2.5) can be solved for any x position along the rod at any t time. In [1], Fourier used – not surprisingly – the Fourier method, and found that the solution for the $T(x, t)$ at any $P(t)$ excitation is a sum of trigonometric functions on x multiplied by exponential functions on t.[8]

[8] An alternative powerful methodology for solving differential equations of this kind is using the Laplace transform. This technique transforms the time-dependent equations into the s complex frequency domain, finds their solution there, and transforms the result back to time domain. Notably, the Fourier and Laplace transforms convert not only functions but also operations. The convolution of two functions in time domain is converted into a simple product of the transformed functions. Calculation of a temperature response on a complicated $P(t)$ function in the convolution integral of (2.12) can be effectively simplified by converting the $g(t)$ and $P(t)$ functions into corresponding $g'(s)$ and $P'(s)$ functions in the complex frequency domain, and then converting their $g'(s) \cdot P'(s)$ product back to time domain. A useful table of Laplace transformed forms of a bunch of functions can be found in [4].

The procedure of the Foster to Cauer-style conversion of RC circuits is also formulated in the s complex frequency domain in Example 2.1 and in Annex C of the measurement standard [40].

In the actual example, Figs. 2.28 and 2.29 represent a case with steady heat flux at the driving point side of the rod (Neumann boundary condition of heat transfer) and steady temperature on the ambient side (Dirichlet boundary condition of heat transfer). Solving the equation for the driving point, we get the already known sum of exponentials as outlined in (2.16) and (2.17).

A more compact analytical solution can be gained with further simplification of the problem. A result of practical importance was introduced in [18] followed by [62] and is used in [40].

Let us apply a P_0 power step on a rod of infinite length, that is, long enough in the sense that its far end remains at ambient temperature for the intended time of investigation. Solving the heat equation in the complex frequency domain at these boundary conditions [62], obtains for the time dependence of the T_J temperature of the driving point:

$$\Delta T_J(t) = \frac{P_0}{A} \cdot k_{\text{therm}} \cdot \sqrt{t} \tag{2.26}$$

where the k_{therm} is

$$k_{\text{therm}} = \frac{2\sqrt{\alpha}}{\lambda\sqrt{\pi}} = \frac{2}{\sqrt{\pi \cdot c_V \cdot \lambda}} \tag{2.27}$$

The thermal diffusivity α used in (2.27) again was defined as $\alpha = \lambda/c_V$ formerly at Eq. (2.5); it is the measure of thermal inertia. In a material of high thermal diffusivity, heat moves rapidly, and the substance conducts heat quickly relative to its volumetric heat capacity.

In practical constructions where the heat generates in a thin "junction" layer on the top of a block of homogeneous material, and starts spreading in that block, the initial section of the temperature transient can be precisely approximated by a square-root-time function.

When the heat spreading reaches the other end of the homogeneous block, then the temperature change takes another shape.

If the homogeneous block is of d thickness conditions, [62] claims that the square-root rule remains valid for short early times of t_v duration:

$$t_v < \frac{d^2}{2\alpha} \tag{2.28}$$

Several thermal properties of typical materials used in the construction of power devices are listed in Table 2.1 (silicon at 25 °C and 125 °C, copper, solder die attach material).

It has to be noted that very thin layers of high thermal conductivity add a very small portion to the temperature elevation of a laminate composed of layers of several constructional materials. In these cases, the "square root of elapsed propagation time" style elevation of the temperature belonging to the next layer can be observed in measured thermal transients.

Table 2.2 lists the valid duration for the square-root-time approach for different materials and die (block) thickness. In the table typical semiconductor and die attach

Table 2.1 Thermal properties of typical materials in power devices

	λ [W/mK]	c_V [J/m^{3}K]	α [mm^2/s]	k_{therm} [m^{2}K/W$\sqrt{s}$]	k_{therm} [mm^{2}K/W$\sqrt{s}$]
Si@25 °C	125	$1.60 \cdot 10^6$	78.3	$7.99 \cdot 10^{-5}$	79.9
Si@125 °C	100	$1.60 \cdot 10^6$	62.7	$8.93 \cdot 10^{-5}$	89.3
Die attach	70	$1.66 \cdot 10^6$	42.1	$1.05 \cdot 10^{-4}$	104.6
Cu	390	$3.40 \cdot 10^6$	114.9	$3.10 \cdot 10^{-5}$	31.0

Table 2.2 Valid duration t_v values for the square-root-time approach

	Si@25 °C	Si@125 °C	Die attach	Cu
d [mm]	t_v [ms]			
0.1	0.063	0.056	0.048	0.161
0.3	0.56	0.50	0.43	1.45
0.5	1.56	1.40	1.20	4.03
1				16.1
2				64
4				258

layer thickness values are shown. Copper is used in very thin layers on printed boards and direct bonded copper (DBC) constructions, and also as bulk material in cold plates. The table helps assigning the subsequent homogeneous spreading regions which can be observed in measured transients to material layers, based on the time range where the square-root-type temperature change occurs.

Example 2.7: Structure Functions of a Real Device

In Fig. 2.30 structure functions of a MOSFET device on a cold plate are shown. This assembly has been used in the former sections as an example for a distributed thermal system.

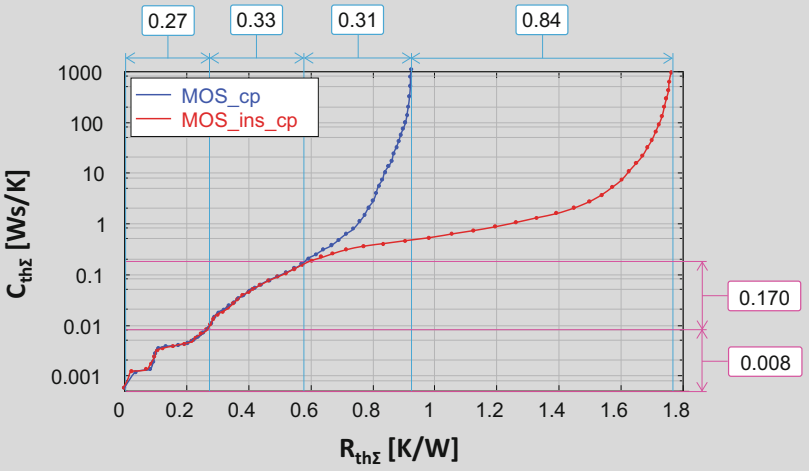

Fig. 2.30 Structure functions of a real distributed parameter system (MOSFET on cold plate, different TIM qualities) with characteristic R_{th} and C_{th} values

(continued)

Example 2.7 (continued)

Curve `MOS_cp` was derived from a thermal transient test when the MOSFET device was mounted on a water-cooled cold plate wetted by a high-quality thermal paste. From the cooling curve (Fig. 2.9), the NID methodology produced 160 RC stages in 1000 iteration steps; these are represented in the time constant spectrum of Fig. 2.19 in a quasi-continuous manner, without displaying each τ value separately. The Foster → Cauer transformation converted these into other 160 RC stages, for which the first 100 are shown as blue dots in Fig. 2.30. The remaining 60 stages are in the 1000 J/K to 10^{38} J/K thermal capacitance range and are not displayed because they are not relevant for the actual study.

In order to distinguish between the device and the test environment, the transient measurement was repeated inserting a ceramics sheet of 2.5 mm thickness between the device package and the cold plate. The transient measurement and the subsequent structure function calculation resulted in curve `MOS_ins_cp`.

This comparison of two structure functions of a device measured at different boundary conditions can be used for deriving standard thermal metrics, as expounded in Sect. 3.1.2 and standardized in [40]. A deeper analysis of structural details which can be recognized in the structure functions will be given in Sect. 7.1 in Example 7.1.

It can be observed that Z_{th} curves in previous figures do not disclose too much details of the structural composition; practically only the junction to ambient thermal resistance value or with multiple boundary conditions an approximate partial resistance until a divergence point, also called bifurcation point, can be read in them. The reason is that the equivalent thermal RC network of the system behaves as a low-pass filter; the sharp power step at its input is converted into the smoothed bumps of the thermal impedance function. On the other hand, the deconvolution algorithms, which produce the structure functions, are closely related to the image enhancement procedures which recreate lost fine details in a blurred picture.

In structure functions many details can be distinguished along with their partial thermal resistance and capacitance value. Still, it has to be mentioned that the structure function analysis is *not* a fully automated ("black box") technique.

There are three ways to assign actual assembly components to sections in the structure function. These are:

- The manufacturer of the device may know all internal geometries and material parameters. In such a way, a "synthetic" structure function can be built up, for example, superposing slices of material with given thermal resistance and capacitance in a spreadsheet tool, and comparing the measured structure functions to it.
- An approximate model can be built up in a finite element or a finite difference simulation tool, such as [56]. Thermal transients can be simulated in the tool and

structure functions can be composed of those. Geometry and material parameters can be fine-tuned until the simulated and measured structure functions match.
- Measured structure functions can be compared to an already identified "golden device." This technique is advantageous in production control.

In the case of Example 2.7, it was easy to measure the external dimensions of the standard TO-220 package which hosts the semiconductor chip. The size of the chip was determined by sectioning the package after the transient test. The assignment of the parts of the structure function to internal details of the same package is presented in Example 7.1 of Chap. 7, Sect. 7.1.

Without this thorough analysis, some characteristic portions of the assembly can be identified in the plot. The structure functions perfectly coincide until 0.6 K/W thermal resistance and 0.178 J/K thermal capacitance, hinting that until this point the heat propagates within the packaged device and the different TIM quality still did not affect the spreading.

The deeper investigation given in the example proves that the first section of the structure functions until 0.27 K/W and 8 mJ/K can be identified as a small silicon chip in the package and the die attach. The next section with 0.33 K/W partial thermal resistance and 0.17 J/K thermal capacitance can be attributed to the heat spreading in the copper tab of the package.

Beyond the identification of the structural elements within the package and the junction to case thermal resistance, also the thermal conductivity of the ceramics can be calculated from the chart. The inserted sheet with its 2.5 mm thickness added 0.84 K/W to the total junction to ambient thermal resistance. The effective cross-sectional area of the heat spreading was limited to the copper surface of the tab, which was 13 mm × 9 mm. According to (2.6), it follows from these geometrical data that the thermal conductivity of the ceramics is $\lambda = 25$ W/mK, a plausible value for sintered alumina material.

The structure function types introduced so far correspond to a one-dimensional mapping of the change of thermal resistance and thermal capacitance along the heat-conducting path. They depict how these local thermal quantities attributed to a section in an assembly change while advancing in the structure from the junction towards the ambient.

The (cumulative) structure function $C_{\text{th}\Sigma}(R_{\text{th}\Sigma})$ demonstrates the growth of the total cumulated thermal capacitance as a function of the total thermal resistance along the heat-flow path. In an alternative view, the differential structure function $dC_{\text{th}\Sigma}/dR_{\text{th}\Sigma}(R_{\text{th}\Sigma})$ was introduced, representing the change of the ratio of the thermal capacitance and thermal resistance, versus the total thermal resistance.

In both representations the thermal properties of structural elements can be identified; these carry the same information; still certain features are more perceptible in one or other form. For example, material interfaces induce a change of steepness in a cumulative structure function, but in its typical logarithmic portrayal, this change may become less apparent. The differential structure function enlarges these differences as obvious local maxima and minima, forming peaks and valleys. Volumes can be best measured in the cumulative version; material interface

locations are often attributed to inflections in the differential one. In other words, the cumulative structure function gives answers on questions of "how much is what we look for?" the differential function rather identifies "where is what we look for?".

2.4.3 The Local Thermal Resistance Function

Some structural elements in an electronics assembly are of well-defined geometry and highly repeatable material properties. Such components are the semiconductor chips and the metal or ceramics parts of the package; their structure is stable in the manufacturing process and later during their lifetime.

As opposed to the above well-defined structural elements, the layers that connect them, the thermal interface materials (TIMs), may show high scatter. Solders and adhesives can be of different thickness and sometimes also of different structure after manufacturing, because of variations in processing steps and heat treatments. Thermal pastes change also later, depending on applied pressure and temperature fluctuations in the normal use of an assembly.

One can observe that the stable constituents are of higher thermal conductivity, resulting in lower thermal resistance. The TIM layers are often very thin, and they add accordingly a smaller portion to the thermal capacitance of a stack; still they contribute a large thermal resistance.

An important aim of structure testing is to find the location of the critical parts of high variation in the structure and to follow their change. A way to magnify the differences in the R_{th} values along the heat-conducting path can be drawing *local* R_{th} values as function of the cumulated thermal capacitances.

This $R_{\mathrm{th}}(C_{\mathrm{th}\Sigma})$ *local thermal resistance function* is also a graphical representation of the Cauer ladder: on the horizontal axis, the sum of the C_{th} elements and on the vertical axis the next R_{th} element of the chain are shown.

The $C_{\mathrm{th}\Sigma}$ values grow monotonously from the origin of the heat towards the ambient. Due to the steadiness of the stable components and the low share in the total thermal capacitance of the TIM layers, the horizontal axis can be considered to correspond to the geometrical location.

Figure 2.31 presents the local thermal resistance functions of four packaged LED samples, soldered to aluminum starboard and mounted on a temperature-controlled cold plate. Sample `1C_W` has a serious delamination problem, which can be identified as a high thermal resistance peak in the 0.2 mJ/K–30 mJ/K range of the local thermal resistance function.

A further example on the use of the local thermal resistance function is presented in Sect. 7.1, Example 7.3. The die attach delamination problem in Fig. 2.31 is treated in depth in Example 7.4.

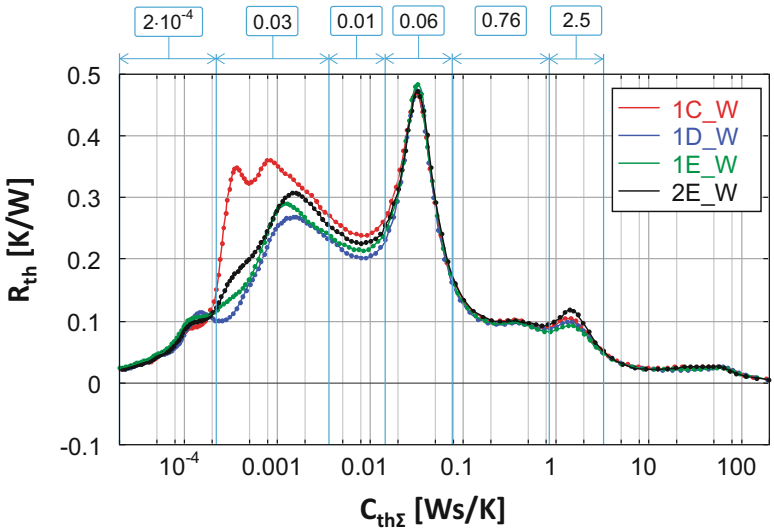

Fig. 2.31 Local thermal resistance functions of packaged LED samples on a temperature-controlled cold plate. Sample 1C_W has a serious delamination problem, as it can be observed in the $C_{th\Sigma} = 0.2$ mJ/K–30 mJ/K range

2.5 Heat-Spreading Patterns in Regular Geometries and Their Appearance in the Structure Functions

In this section we examine some practical cases in which different heat-spreading types can be recognized in the structure functions.

Heat Spreading in a Generalized Tube, Essentially 1D Heat Spreading

Numerical methods can solve the equations of heat spreading for arbitrary shapes and material composition. In the first 150 years of the 200-year history of studying the laws of heat spreading, only analytical methods were available. They can still be used very effectively even today because they yield universal solutions; inserting a few parameters describing the geometry and the material properties, the results are instantly available.

The thermal resistance is always to be measured between two isothermal surfaces in a solid body. In many practical cases, the body can be constructed as a "tube" or "beam," with a surface of varying cross-sectional area shifted along a line or curve (Fig. 2.32).

If the area of the surface is $A(x)$ at position x on the curve, then the thermal resistance of a short dx section is $dR_{th} = (1/\lambda) \cdot (1/A(x)) \cdot dx$, and the added thermal capacitance of the section is $dC_{th} = c_V \cdot A(x) \cdot dx$. Both dR_{th} and dC_{th} are to be added to the cumulated quantities summed up until the point where the dx section started.

Between points x_0 and x_1, the increase of the thermal resistance and thermal capacitance can be calculated as

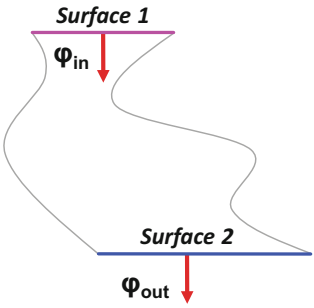

Fig. 2.32 Two isothermal surfaces connected by a heat flux tube

$$R_{\mathrm{th}} = \int dR = \int_{x_0}^{x_1} \frac{1}{\lambda A(x)} dx \quad \text{and} \quad C_{\mathrm{th}} = \int dC = \int_{x_0}^{x_1} c_{\mathrm{V}} A(x) dx \tag{2.29}$$

As suggested by the formulae in (2.29), for the heat transfer in a structure shown in Fig. 2.32, the spatial distribution of both the thermal resistance and the thermal capacitance can be represented as a continuous function of the same independent variable. If this is the case, one can say that the heat spreading in the investigated region is *one dimensional* (see the example of a homogenous rod shown in Figs. 2.27 and 2.28) or *essentially one dimensional*, as discussed below. If the actual dependence of the thermal resistance and thermal capacitance on that common independent variable is known, then the $C_{\mathrm{th}}(R_{\mathrm{th}})$ relationship, i.e., the structure function, is also known, in certain cases given also by analytic formulae.

The Classic 1D Solution: 1D Longitudinal Spreading

As already analyzed in previous sections, the solution of (2.29) is obvious when the area of the isothermal surfaces is constant between the starting and final positions. The object can be considered to be a block or a cylinder, not necessarily a right circular one.

Such a cylinder can be a stand-alone object with a heater on one of its base surfaces, but can also be interpreted as a cylindrical protrusion on a larger object, which emits heat flux into a base surface (Fig. 2.33).

The thermal resistance and capacitance along an L length between x_0 and x_1 positions can be calculated inserting the constant A area as

$$R_{\mathrm{th}}(L) = \frac{1}{\lambda A} \cdot L, \quad C_{\mathrm{th}}(L) = c_{\mathrm{V}} A \cdot L \tag{2.30}$$

Expressing the L length from the first formula and inserting it into the second, we get the analytical expression for the structure function of a cylindrical body:

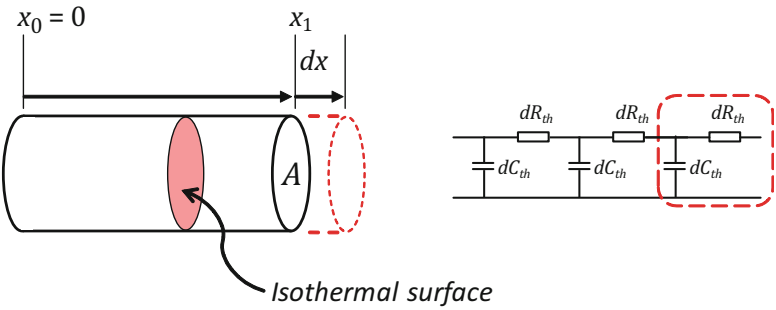

Fig. 2.33 Longitudinal heat spreading in a cylindrical object with constant A isothermal surface area

$$L = \frac{1}{\lambda \cdot A} \cdot R_{\text{th}}, \quad C_{\text{th}}(R_{\text{th}}) = \lambda \cdot c_{\text{V}} \cdot A^2 \cdot R_{\text{th}} \tag{2.31}$$

The differential structure function is $dC_{\text{th}}/dR_{\text{th}} = \lambda \cdot c_{\text{V}} \cdot A^2$ again, but this formula is always valid as previously shown in (2.25).

The Classic 2D Solution: Essentially 1D Radial Spreading

When a smaller heat source is mounted on a heat spreader plate of w thickness, then the heat propagates radially, with concentric isothermal surfaces around the heater position, at least after a certain distance from the position of a heater.

Such a situation can also be represented as a sort of one-dimensional heat spreading, requiring only a transformation of the spatial distribution of the thermal resistance and thermal capacitance to a radial coordinate system. Those cases of heat spreading when a similar transformation from a single space coordinate to another unique space coordinate is possible are considered *essentially one dimensional*.

In a hollow cylinder defined by annular base and w thickness, the dR_{th} and dC_{th} increases can be formulated easily again.

As Fig. 2.34 hints, at radius x the perimeter of a ring is $2\pi x$; thus, in a layer of thickness w, the isothermal lateral surface area around the ring is $A(x) = 2\pi xw$. Adding a thin annular shell of dx infinitesimal thickness to the propagation profile, the lateral thermal resistance growth will be $dR_{\text{th}} = 1/[\lambda \cdot A(x)] \cdot dx$, and the growth in the thermal capacitance when adding an annulus to the existing profile will be $dC_{\text{th}} = c_{\text{V}} \cdot A(x) \cdot dx$.

Integrating between x_0 and x_1 positions, one gets for R_{th}

$$R_{\text{th}} = \int_{x_0}^{x_1} \frac{1}{\lambda \cdot w \cdot 2\pi} \cdot \frac{1}{x} dx = \frac{1}{\lambda \cdot w \cdot 2\pi} \cdot \ln\left(\frac{x_1}{x_0}\right) \tag{2.32}$$

and for C_{th}

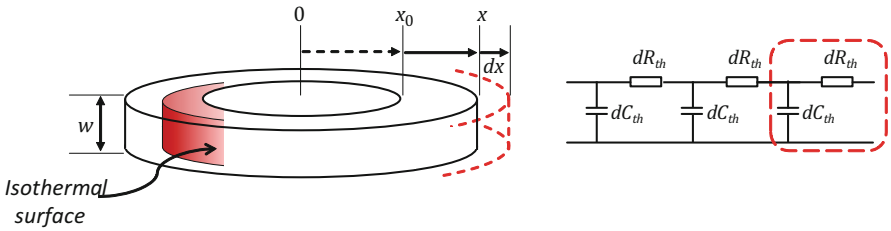

Fig. 2.34 Radial heat spreading in a plate of w thickness

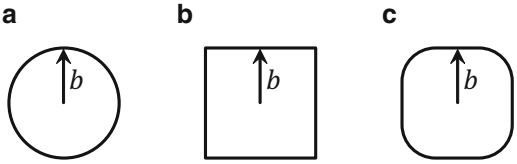

Fig. 2.35 The perimeter of various 2D profiles can be calculated as a characteristic "feature size" b multiplied by an s_p ratio. For the circle (**a**), the ratio is $s_p = 2\pi$, for the square (**b**) $s_p = 4$, while for an arbitrary rounded rectangle (c) such a ratio cannot be given easily

$$C_{th} = \int_{x_0}^{x_1} c_V \cdot w \cdot 2\pi \cdot x \cdot dx = c_V \cdot w \cdot \pi \cdot (x_1^2 - x_0^2) \tag{2.33}$$

In (2.33) one can recognize the area of the annular base as $\pi \cdot (x_1^2 - x_0^2)$. The descriptive formulae for the change of thermal resistance and thermal capacitance are analogous to (2.32) and (2.33) also in other similar cases, when the spreading occurs in a material sheet of w thickness on profiles of *geometrical similarity* along a radial coordinate x in a polar coordinate system. The 2π ratio between the perimeter and the radius of the circle will be replaced by a different s_p factor for other profiles characterized by a b "feature size" (Fig. 2.35).

An example for spreading in circular sectors of growing radius is shown in Chap. 3, Sect. 3.1.2 (Example 3.1 and Figs. 3.3, 3.4, 3.5, 3.6, 3.7, and 3.8).

Expressing x_1 from (2.32) and inserting it into (2.33) in two steps, one gets:

$$\frac{x_1}{x_0} = e^{\lambda \cdot w \cdot 2\pi R_{th}} \tag{2.34}$$

$$C_{th} = c_V \cdot w \cdot \pi \cdot x_0^2 \left(\frac{x_1^2}{x_0^2} - 1 \right) = c_V \cdot w \cdot \pi \cdot x_0^2 \cdot \left(e^{4\pi\lambda \cdot w \cdot R_{th}} - 1 \right) \tag{2.35}$$

The (cumulative) structure function is typically plotted in a lin-log coordinate system as exposed in Sect. 2.4.2.

At larger radii of the heat spreading where $x_1 >> x_0$, the $\ln(C_{th}(R_{th}))$ function will become a straight line, as (2.35) indicates.

Equation (2.35) is of $D \cdot \exp.(4\ \pi\ \lambda\ w\ R_{th})$ form; thus

$$\ln(C_{th}) = \ln(D) + 4\pi\lambda \cdot w \cdot R_{th} \tag{2.36}$$

Taking two points of the structure function plot on a straight section of the $\ln(C_{th}(R_{th}))$ function:

$$\ln(C_{th2}) - \ln(C_{th1}) = \ln(C_{th2}/C_{th1}) = 4\pi\lambda \cdot w \cdot (R_{th2} - R_{th1}) \tag{2.37}$$

and λ can be determined as

$$\lambda \cdot w = \frac{1}{4\pi} \cdot \frac{\ln(C_{th2}/C_{th1})}{(R_{th2} - R_{th1})} \tag{2.38}$$

An example of determining the thermal conductivity of a substrate based on (2.38) is presented in [65].

The Classic 3D Solution: Essentially 1D Conical Spreading

A coordinate transformation can also map a true three-dimensional heat spreading into a corresponding one-dimensional one, if the isothermal surfaces of the spreading conform the principle of *similarity*. In these cases the $A(x)$ surfaces along an x space coordinate grow with a scale factor, or ratio of similarity. For simplicity let us denote the constant scale factor as K.

For determining the scale factor of surfaces in space, first consider a cross section of two typical spreading patterns as shown in Fig. 2.36.

Figure 2.36a corresponds to the spreading in a truncated cone, or truncated pyramid, a standard concept of heat spreading, which is broadly treated in textbooks and also in the literature [60, 65]. In the figure the length of the y vertical sections is always proportional to their distance from the origin, $y(x) = m \cdot x$; the parameter m is the slope of the $y(x)$ function. The area of the flat surface will be $A(x) = y^2\pi = m^2x^2\pi$ if the isothermal surface is circular and $A(x) = 4y^2 = 4m^2x^2$ for a square surface. In

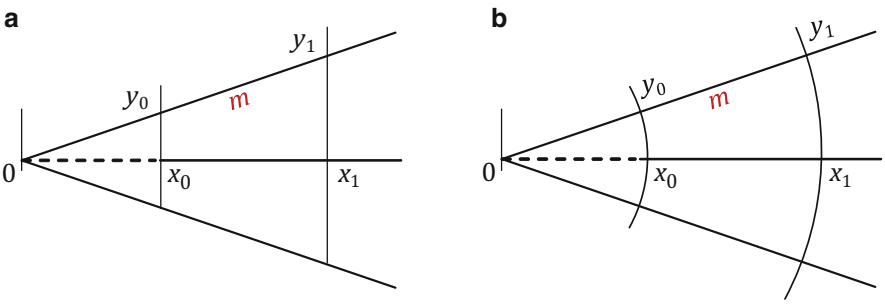

Fig. 2.36 Similar triangles and circular sectors characterized by a 1D growth parameter **m**

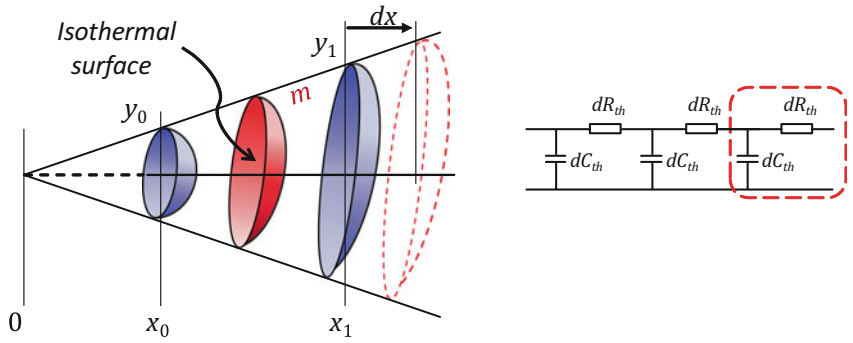

Fig. 2.37 Heat spreading in a "generalized cone," through isothermal surfaces characterized by similarity

both cases the surface grows with an $A(x) = K \cdot x^2$ formula. K can be also interpreted for other shapes, too, like the ones in Fig. 2.35c.

The formulae remain of the same style for spherical spreading, depicted in Fig. 2.36b. In this case a similar m scale factor determines the ratio of the y arc length and the x radius. In geometry m is called central angle, expressed in radians.

For a full sphere, $y = m \cdot x = 2\pi x$ and $A(x) = 4x^2\pi$. For an $m < 2\pi$ central angle value, the surface of the spreading is proportionally smaller:

$$A(x) = 4x^2\pi \cdot \left(\frac{m}{2\pi}\right)^2 = \frac{m^2}{\pi}x^2 \tag{2.39}$$

which is a $K \cdot x^2$ formula again. Analogous formulae depict the growth of the isothermal surfaces of arbitrary shape as illustrated in Fig. 2.37.

The integral between x_0 and x_1 positions yields R_{th} as

$$R_{\text{th}} = \int_{x_0}^{x_1} \frac{1}{\lambda \cdot K} \cdot \frac{1}{x^2} dx = \frac{1}{\lambda \cdot K} \cdot \left(\frac{1}{x_0} - \frac{1}{x_1}\right) \tag{2.40}$$

and C_{th} as

$$C_{\text{th}} = \int_{x_0}^{x_1} c_{\text{V}} \cdot K \cdot x^2 \cdot dx = \frac{c_{\text{V}} \cdot K}{3} \cdot (x_1^3 - x_0^3) \tag{2.41}$$

It can be observed that this generalized "conical" spreading scheme corresponds to several realistic heat propagation patterns, including cones and pyramids.

Selecting the appropriate K scale factor the heat spreading can be calculated in a dome, starting from a small spot in an infinite half space, which is a valid approximation for chips with a hot spot, or bulky heat sinks farther from the mounted device. Similarly, in a larger distance from the actual investigated thermal system, the spreading in the ambient can be considered spherical, obeying (2.40) and (2.41).

The "factor of 1/3" in (2.41) is justified by geometrical considerations, the volume of a cone or pyramid is a third of the enclosing cylinder or block, and the volume of the dome or sphere is two thirds of the circumscribed cylinder. These ratios can be well recognized in actual measured thermal capacitances in real structure function.

In a 3D spreading, the "actual position to ambient" thermal resistance is of finite value; (2.40) yields for "infinite" conical and spherical spreading, from position x_0 towards the ambient $R_{\text{thJA}} = 1/(\lambda K x_0)$. For truncated cones and pyramids, and for spherical shells, (2.40) provides the textbook formula of

$$R_{\text{th}}(L) = \frac{1}{\lambda \cdot K \cdot b_0 \cdot b_1} \cdot L, \tag{2.42}$$

where b_0 and b_1 are the measured "feature size" at the beginning and at the end of the generalized "cone." As defined in Fig. 2.35, b is the radius of a circular shape and half of the edge for a rectangle; K is π for the circle and 4 for the rectangle. It can be seen that (2.42) describes the heat propagation over growing surfaces along an x coordinate as a "generalized rod" where the surface in (2.30) is replaced by the product of the linear b_0, b_1 dimensions at the beginning and at the end. Still, the junction to ambient thermal resistance remains finite, as the L/b_1 ratio converges to the constant m factor for large lengths.

Further research results on the spreading shapes are published in [66, 67, 139].

As illustrated in Fig. 2.38, in real package structures, one often finds actual heat-flow paths with different sections that can be characterized with 1D or essentially 1D heat-spreading patterns discussed above.

In summary, we can state that in many cases, there is a one-to-one correspondence between the physical sections of the junction-to-ambient heat-flow path and the structure functions. There are cases, however, where there is a real, complex heat-spreading pattern within a package where one cannot identify any dominant heat-flow path. In such a situation, the structure function (as one of the representations of a junction-to-ambient thermal impedance) is the *thermal signature* of the

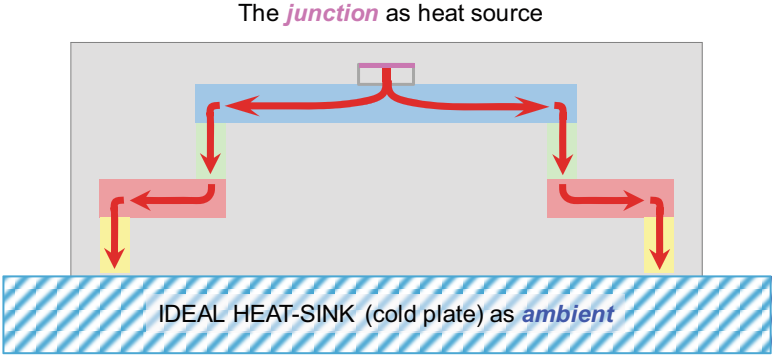

Fig. 2.38 Real package structures often can be represented as a series of different heat-flow path sections characterized by (essentially) 1D heat spreading

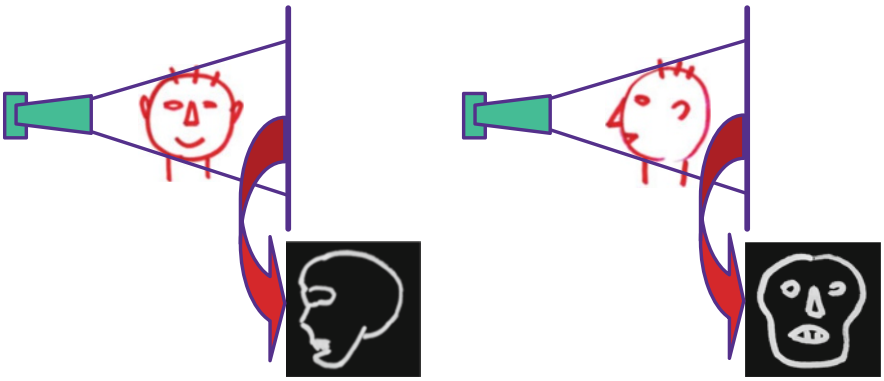

Fig. 2.39 Structure functions can be considered as "1D projections" of complex 3D heat-spreading patterns like X-ray images are 2D projections of three-dimensional bodies

system; it can be considered only as a one-dimensional "projection" of the complex 3D pattern. Note that this signature is still perfectly applicable to test the integrity of the structure.

Measuring the thermal impedance of such packages on cold plates helps un-blurring the image, though the heat is directed mainly into one major path, towards that surface of the package that is in direct contact with the cold plate during the test, while all other surfaces of the package are thermally isolated. Changing the test setup such that another surface of the package is in contact with the cold plate while the previous one is thermally isolated from the ambient directs most of the heat into another path.

The structure functions obtained this way are somewhat similar to one's taking X-ray images from different directions, as illustrated in Fig. 2.39. Thermal "CAT-scan images," though, cannot be constructed from measured junction temperature transients.

The DELPHI methodology [68] defines four different boundary conditions with various high and low heat transfer coefficients on different package surfaces for validating simulated package models by measurements. The actual boundary conditions are presented in Chap. 7, Sect. 7.7.3.

2.6 The Concept of the Heat Transfer Coefficient

In the previous section, the thermal properties of structural elements were formulated in Eqs. (2.30)–(2.42), and more generally in (2.29). The thermal resistances were calculated assuming that the geometry of the heat-conducting path is known in details.

Table 2.3 Typical values of convective heat transfer coefficient

Convection type	h [W/m^{2}K]
Air, free	2.5–25
Air, forced	10–500
Liquids, forced	100–15,000
Boiling water	2500–25,000
Condensing water vapor	5000–100,000
Surface of a cold plate	1000–5000

In many cases it is more practical to use the concept of the h heat transfer coefficient, the proportionality factor between the φ heat flux through and the ΔT temperature between two isothermal surfaces.

The h coefficient is in common use as a thermal figure of merit of commodities manufactured in specific thickness and composition. It is widely used for TIM sheets or coatings in electronics or similarly for characterizing thermal insulation or glazing in building industry.[9]

Another field of use is to characterize the termination of the heat-conducting path in an assembly. In case of convective cooling, many formulae are used for determining the cooling capability of a gas or fluid flow on a solid surface. The formulae are mostly empirical and take into consideration the nature of the fluid flow (natural, laminar, turbulent flow) and material-related coefficients (Prandtl, Rayleigh, Nusselt numbers, etc.). Table 2.3 lists approximate heat transfer coefficients for some convective cooling solutions used in power electronics. Boiling water and vapor condensation were included in the table because of the growing importance of heat pipes in electronics cooling.

As the heat transfer coefficient is defined as $h = \varphi/\Delta T$, the thermal resistance of an inserted material or a heat exchanger of A surface can be calculated as

$$R_{\text{th}} = \frac{1}{h \cdot A} \tag{2.43}$$

2.7 Driving Point and Transfer Impedances: Self-Heating and Transfer Heating

So far only such cases were investigated where a single point was heated, its own temperature was measured, and the whole system was reduced into a single thermal RC network model. There are many practical situations, however, where (due to different reasons not discussed here) one cannot measure the temperature response at the same location where the heating was applied. Typical examples are laterally arranged *multi-die packages* (such as illustrated in Fig. 2.40), stacked die packages,

[9] The heat transfer coefficient is denoted as U or K factor in the building industry.

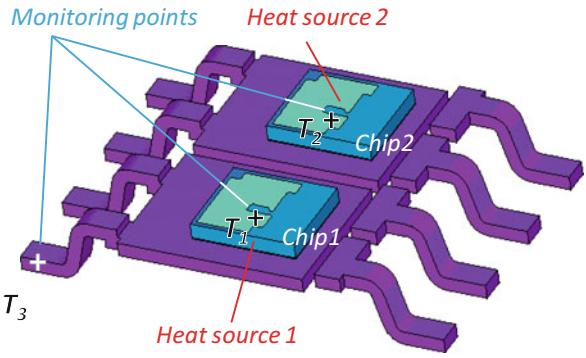

Fig. 2.40 Axonometric view of the major parts of a package of a dual-chip power device with two heat sources (driving points) and three temperature monitoring points. T_1, T_2, and T_3 represent the temperatures of the two chips and the temperature of one of the pins of the device, respectively

thermal test chips [32], LED packages with multiple LED chips, or PCB-assembled LED modules with a thermal test point on the board. In case of a multi-die package or an RGB LED module one is interested, how the heat dissipation of one semiconductor chip affects the temperature of the other dies in the system.

To allow a proper distinction and precise description of such cases, the concepts of *driving points* and *temperature monitoring points* were introduced in the technical literature [69].

The *driving points* are the *locations of the heat sources in the system* (i.e., the junctions of the chips).

The *temperature monitoring points* are the *locations where the temperature responses are measured*.

In most practical cases, as in the examples discussed previously, the *junctions* of the chips are both driving points and temperature monitoring points. The thermal impedance obtained from measurements when the power step is applied and the temperature response is captured at the same location is called *driving point thermal impedance* or *self-impedance* in short.

When the driving point and the temperature monitoring point are separated in space, the thermal impedance obtained is called *thermal transfer impedance*, in this book also referred to shortly as *transfer impedance*.

Example 2.8: Self-Heating and Transfer Heating in a DDR RAM Module

As an example for the driving point and transfer impedances, a DDR RAM module, mounted into the socket of a PC motherboard, was simulated. Applying 0.2 W power on the internal device `Chip1` in the leftmost RAM package `RAM_1`, the temperature distribution, which develops on the module surface in steady state, is shown in Fig. 2.41.

(continued)

Example 2.8 (continued)

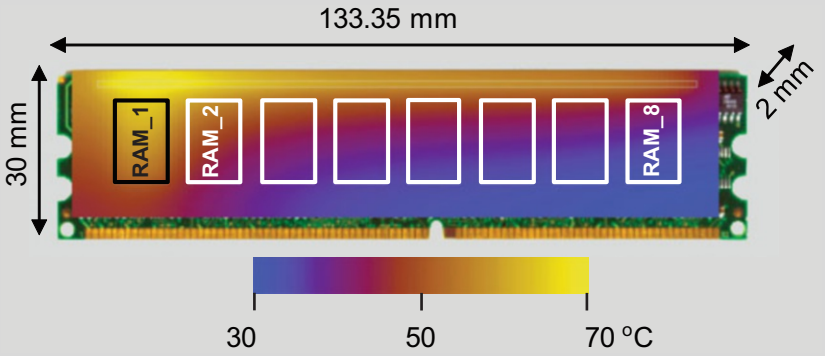

Fig. 2.41 Simulated temperature distribution in a DDR RAM module in PC socket, 0.2 W applied on the leftmost RAM package `RAM_1`

As a further example, Fig. 2.42 presents the simulated transients on the internal chips in `Chip1`, `Chip2`, and `Chip8` encapsulated into packages `RAM_1`, `RAM_2`, and `RAM_8`. The self-heating curve of the driving point `Chip1` starts growing at early time, a delay of 1.1 s can be observed in the transfer curve of `Chip2`, and it takes nearly 10s until the heat propagates towards monitoring point `Chip8`.

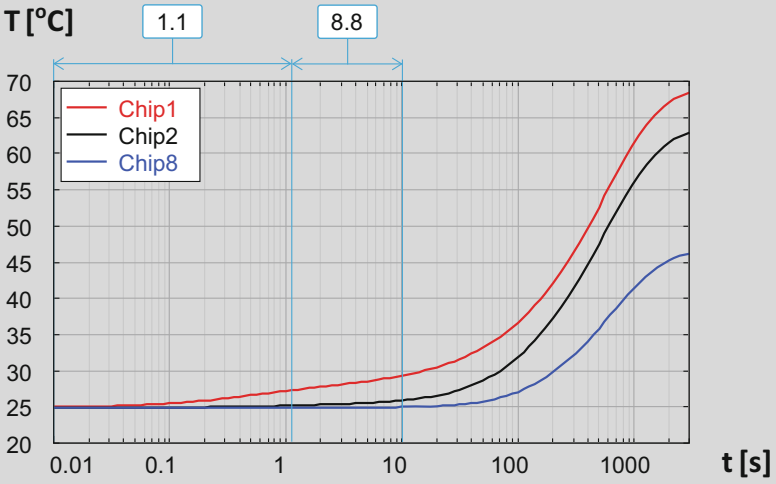

Fig. 2.42 Simulated temperature response of the DDR RAM module in PC socket, 0.2 W applied on the leftmost RAM package. Transients at the driving point Chip1 and monitoring points Chip2 and Chip8 are shown

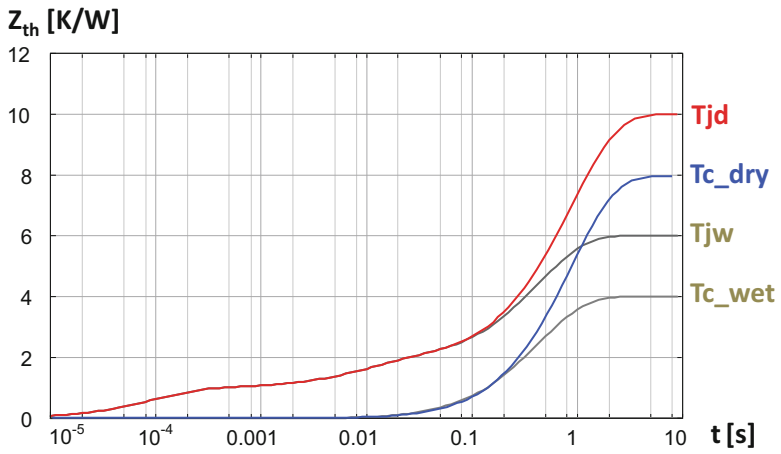

Fig. 2.43 Self-impedance curves `Tjd` and `Tjw` and transfer curves `Tc_dry` and `Tc_wet` of the two simple Cauer models represent the dry and wet boundary conditions in Fig. 2.21 of Example 2.5. The naming of the curves corresponds to the node names in the model

Physical considerations imply that the temperature response at a *single driving point* is always monotonous; the temperature constantly grows when power is applied and decreases when it is revoked. Similarly, also transfer curves remain monotonous if the structure is nearly one dimensional between the driving point and the monitoring point.

The simulation of the two simple Cauer-style models in Fig. 2.21 of Example 2.5 yields the Z_{th} curves of Fig. 2.43. The curves denoted with `Tjd` and `Tc_dry` correspond to the simulated self-impedance and transfer impedance of the model at dry boundary condition, represented by higher thermal resistance from the internal `Tc_dry` location towards the ambient. The curves `Tjw` and `Tc_wet` show the results for the wet boundary condition.

The *propagation delay* between the driving and the monitoring points is characteristic to the transfer impedances; it is approximately 10 ms in this assembly.

The propagation delay in the time domain curves of the transfer impedances results in some negative magnitude values in the time constant spectra.

The monotonous nature of the transfer curves cannot be assured in systems with multiple heat sources or multiple heat-conducting paths within the thermal system. The root cause of this non-monotonous behavior and its practical conscequences will be treated in detail in Sect. 3.5.

A thermal system with multiple driving points and temperature monitoring points (such as a multi-die system; see Fig. 2.44) can be fully represented if all its possible self- and transfer impedances are known, i.e., all of these thermal impedances are measured or simulated.

On the right side of Fig. 2.44, the Z^*_{ii} and Z^*_{jk} symbols representing all possible thermal impedances are arranged in a matrix. In [69] such matrices are called *thermal transfer impedance matrices*. This is the dynamic extension of the concept

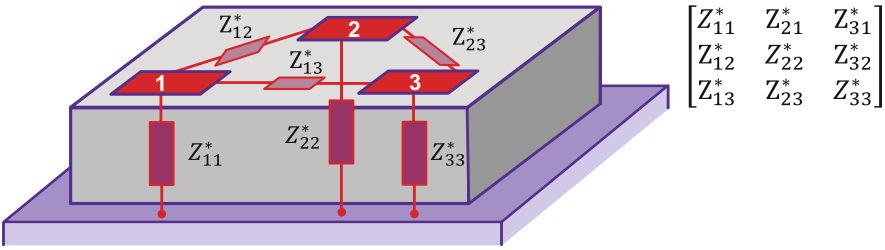

Fig. 2.44 A substrate with three heat sources (i.e., driving points) that are also used as temperature monitoring points and the illustration of the corresponding self- and transfer impedances

of the *thermal resistance matrix* introduced among others in [70]. Nowadays such a matrix is called *thermal characterization matrix* when steady-state values are included in the matrix (usually denoted by $\boldsymbol{R}^*_{\mathbf{th}}$) or *dynamic thermal characterization matrix* (usually denoted by $\boldsymbol{Z}^*_{\mathbf{th}}$) when thermal impedances are therein. The features and issues related to these thermal characterization matrices will be discussed later in Chap. 3 dealing with the so-called thermal metrics.

Note that the elements included in these thermal characterization matrices do not represent the element values of a so-called multi-port compact thermal model of a multi-die system. Steps of obtaining a steady-state thermal model from the $\boldsymbol{R}^*_{\mathbf{th}}$ matrix are described, e.g., in [71]; an application for dynamic compact thermal modeling of digital IC chips based on the $\boldsymbol{Z}^*_{\mathbf{th}}$ matrix is described in [72].

2.8 System Descriptors for Periodic Excitations

Power electronics applications are mostly exposed to periodic excitations. Some appliances, such as motors, generators, or the input side of power supplies, are directly connected to the power grid, which operates at sinusoid alternating current. Other applications, such as car and locomotive electronic traction control units (ECUs), PWM controls in LED lighting, and internal circuitries of switching power supplies, operate at pulsed direct current.

It was stated in the previous sections that the linear network theory enables the calculation of the system response on arbitrary excitations from the measured step response. The methodologies offered by the linear approach are particularly suitable for directly deriving system responses on periodic excitations.

Below two cases of practical importance are discussed. First, the concept of the *complex locus*, a system descriptor that enables the direct production of the response of a thermal system on arbitrary periodic power change, will be presented. The methodology expects the spectral decomposition, that is, the frequency, amplitude, and phase of the constituents of the power as input, and yields the temperature response in the frequency domain. The apparatus of the Fourier transform and inverse Fourier transform connects the excitation and the response to their view in

time domain. The technique is equally suitable for calculating the response of driving and monitoring points.

The use of complex loci is indispensable in the investigation of the stability of systems with thermal feedback, for tuning the control loop of thermostats at a given thermal mass, etc.

The concept of *pulse thermal resistance diagrams* has a simpler and less universal use. The thermal system is represented as a parametrized set of curves from which the peak temperature in the stationary state can be read when an excitation of repeated pulses of known period time and duty cycle is applied.

2.8.1 Complex Loci

The frequency domain representation of the thermal impedance can be calculated from the time domain $Z_{\text{th}}(t)$ function. The Fourier transform yields the $Z_{\text{th}}(\omega)$ function as

$$Z_{\text{th}}(\omega) = \int_0^\infty Z_{\text{th}}(t) \cdot e^{-j\omega t} dt \tag{2.44}$$

where ω is the angular frequency of the excitation. The resulting $Z_{\text{th}}(\omega)$ complex thermal impedance function can be visualized, e.g., by means of a *complex locus*, also known as Nyquist diagram.

In Example 2.5 an artificial Cauer model consisting of three RC ladder stages was presented; the time domain Z_{th} curves characterizing the self-impedance and transfer impedance were shown in Fig. 2.43.

Applying (2.44) on the self-impedance curves `Tjd` and `Tjw` corresponding to the "dry" and "wet" boundary condition, the complex loci of Fig. 2.45 can be gained.

The complex loci depict how a thermal system responds to a unit-size sinusoid power of the f frequency and the $2\pi f$ angular frequency.[10] In the chart, the Re real part of the thermal impedance corresponds to dissipation, and the Im imaginary part expresses heat storage. The amplitude of the temperature response is the length (absolute value) of the vector between the origin and the corresponding point in the plot; the phase shift between the power and the temperature is the angle between the x axis and the vector, representing the delay in the temperature response. As the temperature always lags behind the power, the angle is negative; consequently the thermal impedance curves belonging to a driving point are in the fourth quadrant of the complex plane.

At a single sinusoid excitation, the thermal system behaves as a thermal resistance of Re $Z_{\text{th}}(\omega)$ and a thermal capacitance calculated from the formula Im $Z_{\text{th}}(\omega) = 1/$

[10] The linear network theory also uses an alternative representation of the same information. The Bode diagrams display the absolute value and the phase of the system response in separated charts.

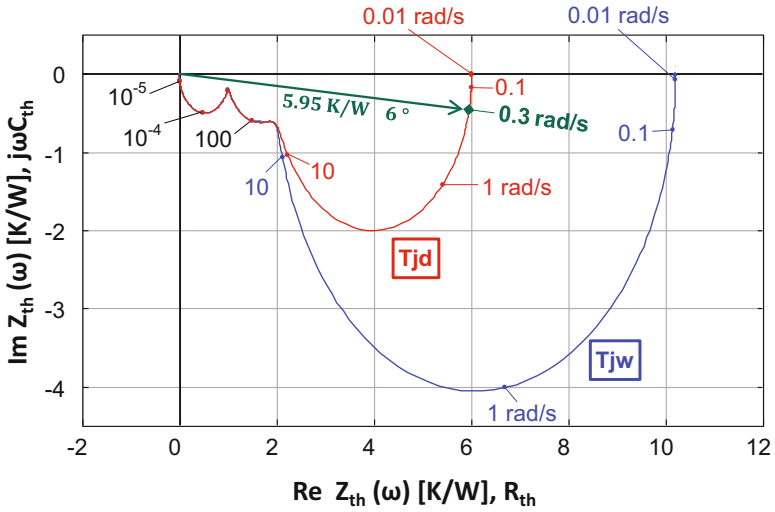

Fig. 2.45 Complex loci calculated from the time domain Z_{th} curves of Fig. 2.43, characterizing the self-impedance in two simple systems of different boundary conditions, as presented in Example 2.5

($j\omega C_{th}$). The physical meaning of the smaller temperature change on the same power amplitude is that the periodically changing heat can be locally stored and released by the structures near to the junction; it does not reach the ambient.

A single parallel RC circuit has the locus of a half circle, as its points can be calculated from the fixed R_{th} thermal resistance value and the shrinking $j\omega C_{th}$ thermal admittance. In Fig. 2.45 the portions of the three half circles can be clearly recognized showing the three discrete time constants of the system.

The zero-frequency value expresses the R_{thJA} junction to ambient thermal resistance at the two boundary conditions. When a periodic power signal is applied on the thermal system, which can be decomposed into several single frequency components of various amplitude and phase angles, the system responds in the frequency domain with a temperature response composed as the sum of the $Z_{th}(\omega)$ vectors.

Figure 2.46 shows the self-impedance and transfer impedance, calculated from the time domain Z_{th} curves `Tjd` and `Tc_dry` in Fig. 2.43, belonging to the "dry" boundary. In the enlarged excerpt of Fig. 2.47, it can be observed that the temperature change at the `Tc_dry` monitoring point is of a phase shift higher than 90 ° at frequencies above 40 rad/s which corresponds to 12 Hz. In general, complex loci corresponding to transfer impedances extend both to the third and fourth quarters.

This can be also interpreted in such a way that because of the propagation delay between the points `Tjd` and `Tc_dry`, the temperature change can be in an opposite phase as compared to the power excitation.

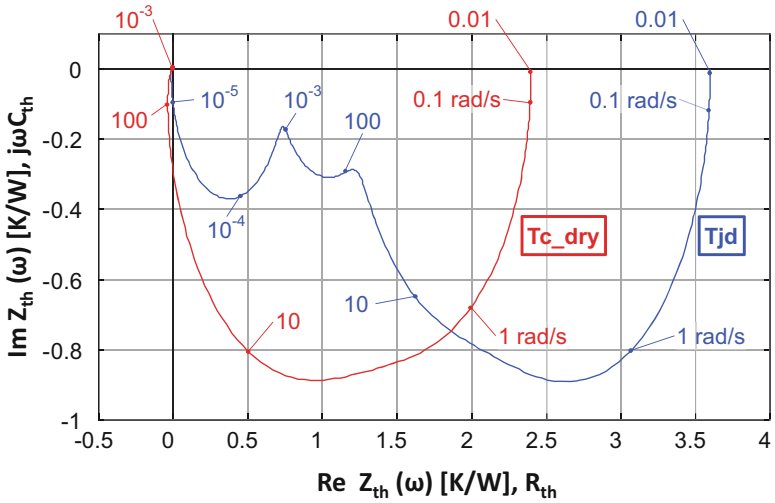

Fig. 2.46 Complex loci calculated from the time domain Z_{th} curves in Fig. 2.43 belonging to the "dry" boundary, characterizing self-heating (`Tjd`) and transfer heating (`Tc_dry`)

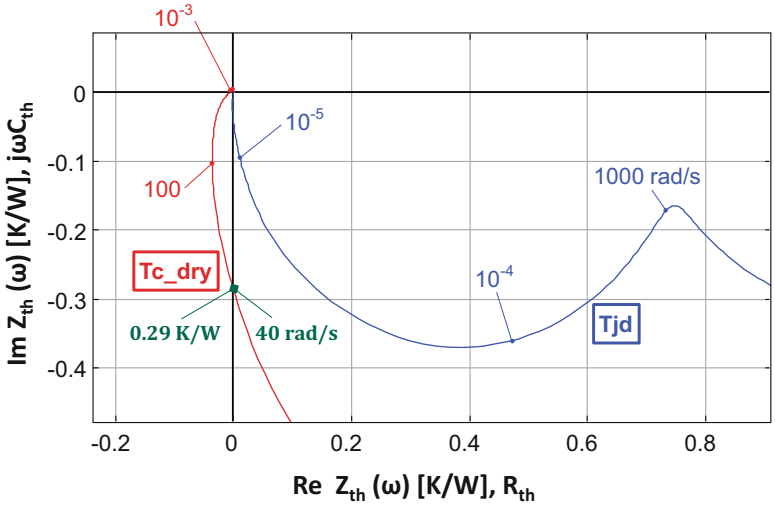

Fig. 2.47 Excerpt of Fig. 2.45. With positive feedback and high enough gain, thermal oscillation can be expected above ~40 rad/s corresponding to ~12 Hz

In this system if the coupling between the electric side of the powering and the `Tc_dry` point has a high enough gain, the positive feedback can induce a thermal oscillation above 12 Hz.

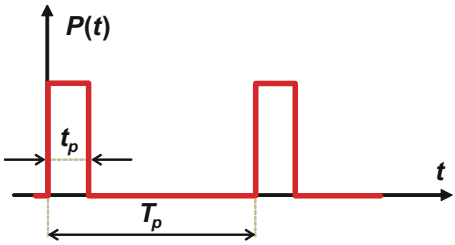

Fig. 2.48 Periodic power pulse sequence

Complex loci are a very powerful representation of the component and its environment when analyzing periodic excitations. An application example is the single valued "AC thermal impedance" of LEDs [73], detailed in Sect. 6.10.

2.8.2 *Pulse Thermal Resistance Diagrams*

So far we have discussed the thermal characterization of a system based on its response to a single power step. In an important class of practical applications such as switching power supplies and motor drives, the power excitation can be described as a series of repeated power pulses. Similar pulse sequences play an important role, e.g., in reliability testing (Chap. 7, Sect. 7.4).

A periodic pulse sequence can be characterized by the T_p period time and the t_p length of the "on" state (Fig. 2.48), or by the T_p period time and the $D = t_\mathrm{p}/T_\mathrm{p}$ duty cycle.

With an excitation of repetitive pulses at a certain duty cycle, some heat will be stored in the internal thermal capacitances during the "on" state of the pulse and will be released during the "off" state. In a more detailed view, if the pulse is applied at a certain location of the system, the thermal energy is first stored in the material sections in close vicinity of the excitation, and farther sections are filled up with thermal energy at longer times as the heat propagates.

If T_p is longer than the shortest relevant time constant of the system, the thermal capacitances cannot be fully discharged in the "off" state; the average temperature continuously elevates in the system until stationary state is reached.

Example 2.9: Momentary Temperature Change and Peak Temperature at Pulse Excitation

Two versions of a simplified heat-conducting path were presented in Fig. 2.21, representing a DUT on a cooling mount. Different thermal interface qualities were denoted as "wet" and "dry" boundary condition.

In the following example, the "dry" scheme was driven in three subsequent simulations by different pulse sequences (Fig. 2.49).

(continued)

Example 2.9 (continued)

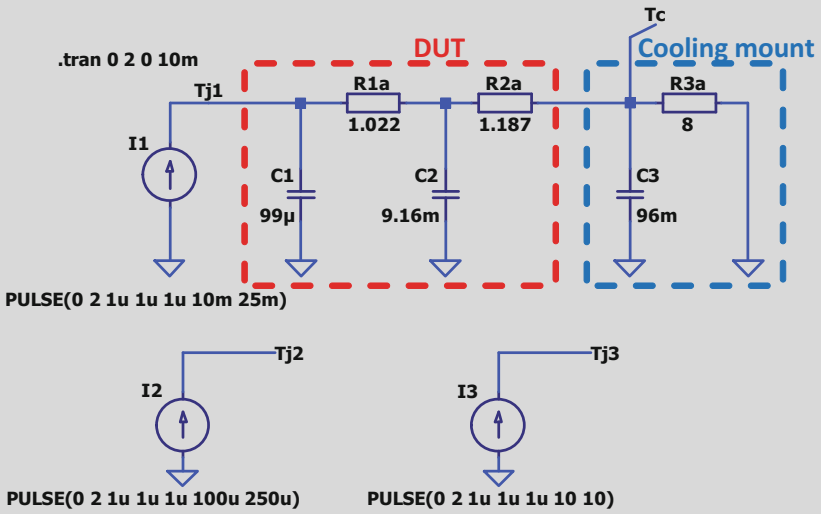

Fig. 2.49 Equivalent RC ladder scheme of a heat-conducting path, corresponding to the "dry" boundary condition in Fig. 2.21, pulsed power excitation at different period time and duty cycle

The transients were simulated until 2 s. In the first simulation, the circuit was driven by 2 W power, with 40% duty cycle, 25 ms period time, and 10 ms "on" time. The result of the simulation is shown as the red `Tj1` line in Fig. 2.50. The second simulation was carried out with the same duty cycle but with 250 µs period time (`Tj2`, green line). A "long" step excitation was applied in the third simulation, at least longer than the total 2 s simulation time.

It can be observed in Fig. 2.50 that, as a consequence of the linear approach, the average temperature of the transients taken at 40% duty cycle is around 40% of the transient temperature in the step response. However, for thermal overload and lifetime prediction, the *peak temperature* reached at pulsed powering is of primary interest.

This peak temperature is highest in a pulsed heating process when stationary state is reached. This occurs when the elapsed heating time exceeds a few times the largest time constant of the system; in Fig. 2.50 this is reached at a few seconds.

(continued)

Example 2.9 (continued)

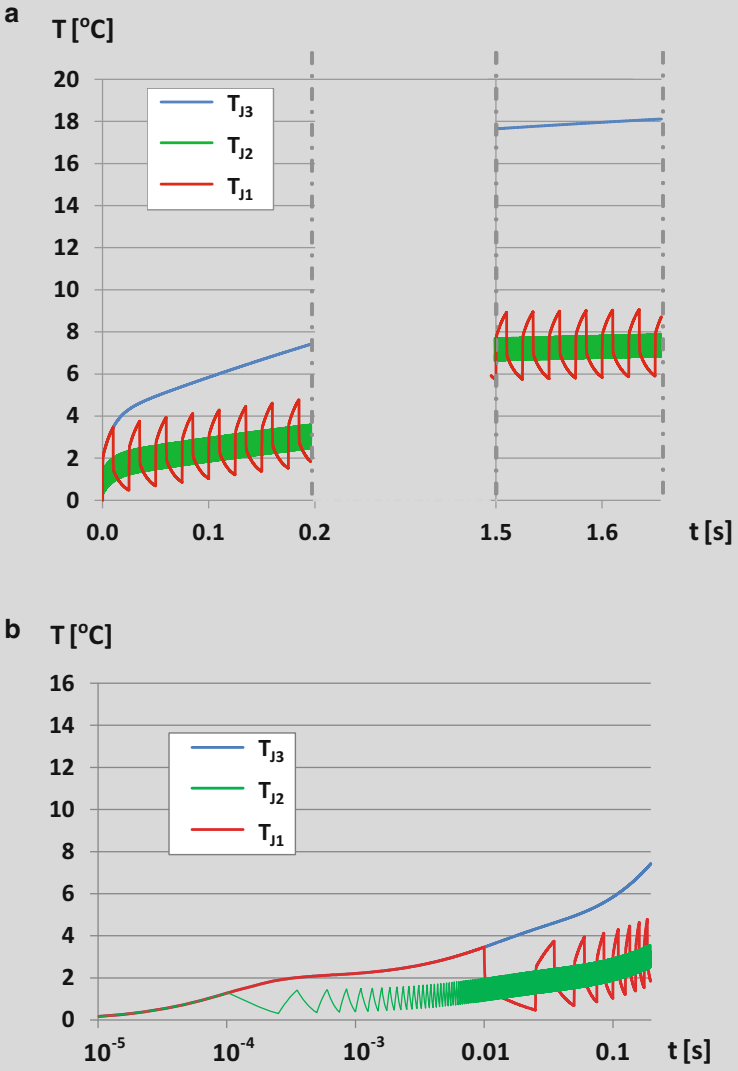

Fig. 2.50 Simulated results with pulsed power excitation at 40% duty cycle and different period time: `Tj1`, $T_p = 25$ ms; `Tj2`, $T_p = 250$ µs; `Tj3`, step excitation (**a**) linear time scale (**b**) logarithmic time scale

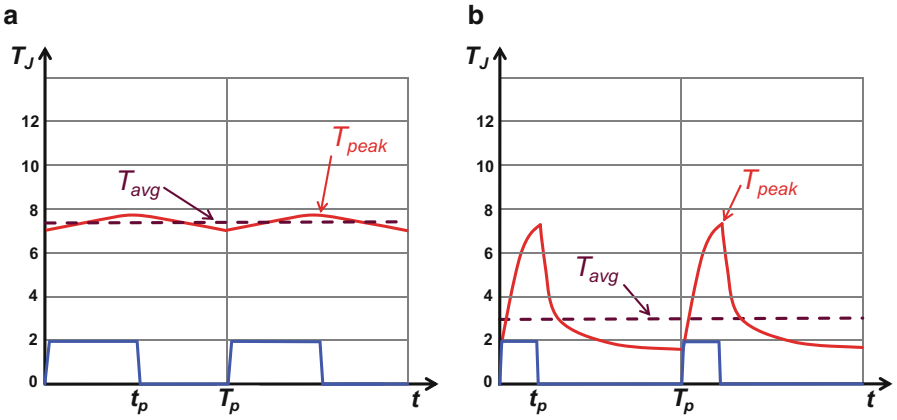

Fig. 2.51 Junction temperature waveform at different heating pulse series: (**a**) $t_p = 0.1$ ms, $T_p = 0.2$ ms, $D = 0.5$, $f = 1/T_p = 5$ kHz, $dT_{peak} = 7.8$ K, $dT_{avg} = 7.3$ K. (**b**) $t_p = 1$ ms, $T_p = 5$ ms, $D = 0.2$, $f = 1/T_p = 200$ Hz, $dT_{peak} = 7.9$ K, $dT_{avg} = 2.9$ K

In stationary state the junction temperature follows a periodic function of stable waveform, similar to the ones shown in Fig. 2.51.

In an actual thermal testing of a real system, one way to establish the peak value of the periodic temperature response is applying different pulses of different length and amplitude and measuring the peak temperature directly. For practical reasons this can be done in the "off" state just after the falling edge of the power pulse as suggested in Sect. 5.4.1.

Carrying out a series of measurements with square wave excitation at many frequencies and duty cycles is rather tedious. Instead, a plot representing the peak temperature at several T period times and D duty cycles can be derived from a single measurement with a sole step function excitation, followed by some mathematical calculations. This plot, called *pulse thermal impedance* plot, can be calculated from the time constant spectrum.

The actual temperature waveform can be determined by measurement and simulation. First, one of the RC models of the device defined in previous sections is to be constructed from a thermal test with step excitation, and then a SPICE-like circuit simulator can produce the waveforms. This technique is often used for creating the pulse thermal resistance diagrams for data sheets; however, a direct methodology based on LTI theory can produce the plot in a single convolution step.

Equation (2.21) yields the $R_\zeta(z)$ time constant spectrum from a measured step response. In a linear approach, a single power pulse can be interpreted as a pair of consecutive step functions; a negative power step after t_p time extinguishes the first positive one. The corresponding thermal responses can also be superposed, deferred by t_p time and with opposite sign.

This concept can be extended to a series of pulses which follow each other by a T period time.

The direct calculation of the convolution integral of (2.12) in the time domain would need numerical approximate formulae. However, periodic functions like the pulse sequence discussed so far have a compact form in the (complex) frequency domain. The solution becomes of manageable complexity carrying out the following operations again:

- Transforming the periodic excitation and the (logarithmic) time constant spectrum by Fourier (or Laplace) transform into the (complex) frequency domain
- Multiplying the transformed functions
- Applying the inverse Fourier (or Laplace) transform to get the (logarithmic) time domain solution

In case of a periodic pulse excitation with t_p pulse width, T_p period time and $D = t_p/T_p$ duty factor (Fig. 2.48), the curves of the pulse thermal resistance diagram can be calculated from the time constant spectrum by the following convolution operation:

$$Z_{th}(z_p = \ln\ t_p, D) = R_\zeta(z) \otimes \frac{1 - \exp\ [-\exp\ (z)]}{1 - \exp\ [-\exp\ (z)/D]} \tag{2.45}$$

From the $Z_{th}(z_p,D)$ result, substituting z_p with ln t_p a $Z_{th}(t_p,D)$ function can be constructed again. Keeping the convention of engineering practice used so far, in charts $Z_{th}(z_p,D)$ will be plotted, but the horizontal axes will be labeled with t_p and the plot will be referred to as $Z_{th}(t_p,D)$.

The result of (2.45) is far from being mere theory; it is the mathematical expression of a very practical algorithm which can convert the result of a single transient measurement (test response) into the pulsed thermal impedance, without the need to test a structure by pulse patterns of various t_p, T and D parameters. An example of this calculation is presented in [60]. An actual software tool which realizes the calculation is part of the tester configuration of [54].

The calculated $Z_{th}(t_p,D)$ pulse thermal resistance plots for the scheme of Fig. 2.49 are shown in Fig. 2.52. Several duty cycles are plotted in the 5%–50% range. The curves were distilled from the response of the *whole RC ladder*, including both the DUT part and the cooling mount part. Accordingly, the chart can be called as junction to ambient pulse thermal resistance plot and can be denoted as $Z_{thJA}(t_p,D)$; it reflects the temporary energy storage on both the device and the cooling mount sections. The peak temperature of the stationary state can be calculated for power pulses of P height as $T_{peak} = P \cdot Z_{thJA}(t_p,D)$.

The $D = 0\%$ curve corresponds to the Z_{th} curve of the "dry" boundary scheme, drawn as `3stageCauer_dry` plot in Fig. 2.23. At long period times, the temperature can reach its full $T_{peak} = P \cdot R_{thJA}$ value. At high repetition frequencies, that is at low t_p the peak junction temperature equals approximately the average temperature, $Z_{th} = D \cdot R_{thJA}$ and $T_{peak} = P \cdot D \cdot R_{thJA}$. In Fig. 2.52 the junction to ambient

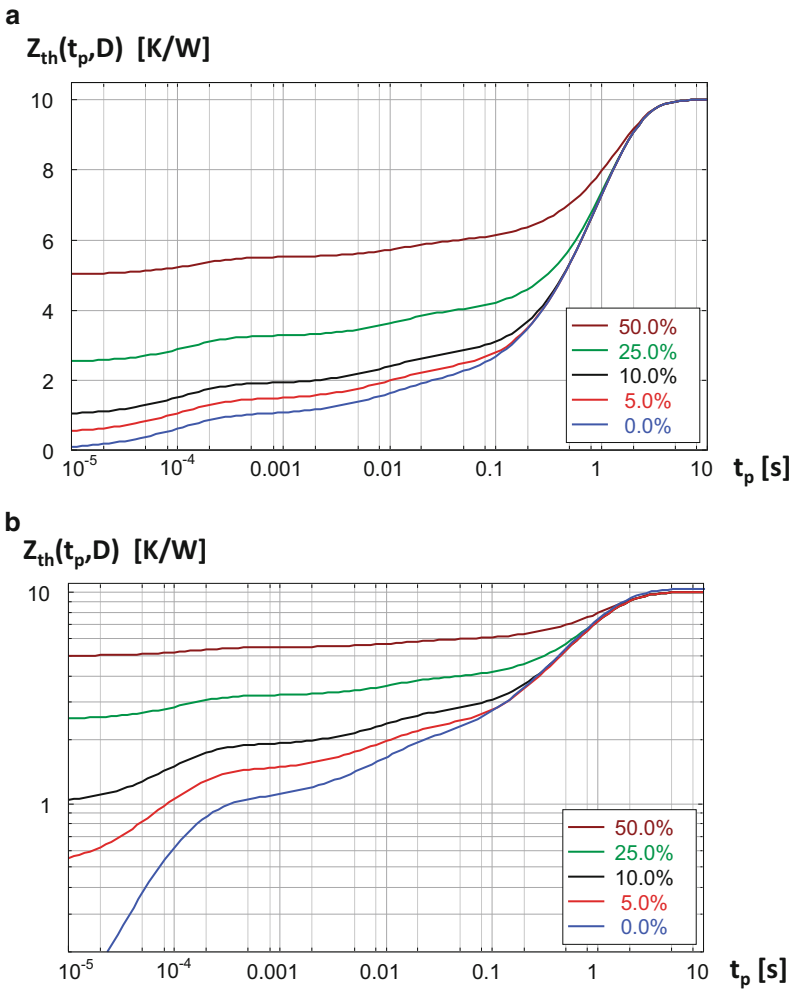

Fig. 2.52 $Z_{\mathrm{thJA}}(t_{\mathrm{p}},D)$ junction to ambient pulse thermal resistance of the heat-conducting path scheme of Fig. 2.49; (**a**) linear, (**b**) logarithmic pulse thermal resistance scale

thermal resistance is $R_{\mathrm{thJA}} = 10$ K/W; we can observe at short pulses the pulse thermal resistance value of 5 K/W, 2.5 K/W, and 1 K/W for duty cycles 50%, 25%, and 10%, respectively.

Data sheets typically present simulated charts in logarithmic pulse thermal resistance scale. These charts are mostly "reduced" to the "case" surface of a packaged device or module, in various, and mostly doubtful ways.

In a more sophisticated technique, a simulation is carried out on the detailed geometry of the device at constant uniform temperature on the baseplate, and the transient results are used to compose the pulse thermal resistance chart.

In a simpler approach, a Cauer RC network is composed from a measured thermal transient of the device in a conductive test environment (defined in Chap. 5, Sect. 5.1), practically on a cold plate. Based on some assumptions, a single internal node in the Cauer ladder is denoted as the "case" and that point is connected to the ambient. An analog circuit simulator software tool is used to compose the pulsed response of the shortened ladder.

In an even less justified approach, a Foster equivalent of a few stages is presented as junction to case thermal model. A seen before, contrary to the Cauer model, when a Foster model is terminated with different thermal impedances, the element values in the chain vary, as the Foster model is only valid for a single boundary condition. In other words, the Cauer model is a boundary condition-independent (BCI) model for the cases, when there is only one heat-flow path from the junction to the ambient, while the Foster model is always a boundary condition-dependent model.

A $Z_{\mathrm{thJC}}(t_{\mathrm{p}},D)$ "junction to case" pulse thermal resistance chart is presented in Fig. 2.53. The plot was composed using the scheme of Fig. 2.49 again; the T_{C} node in the Cauer ladder was assumed to be the interface between the models of the device and the cooling mount. For mimicking the broadly used but theoretically wrong procedure, T_{C} was connected to the ambient (ground), and the pulse thermal resistance was constructed with (2.45). The chart suggests that the heat spreading in the modeled device reaches the assumed "case" surface at approximately 50 ms; the partial thermal resistance between the junction and the case node is 2.2 K/W, as also known from the circuit scheme.

This process for deriving the junction to case pulse thermal resistance is quite ill-defined; in reality no isothermal "case" surface exists, and the complicated trajectories of heat spreading cannot be reduced to a single one-dimensional RC network with a dedicated case node. These ambiguities will be treated more in detail in Chap. 3, where a standard R_{thJC} junction to case thermal resistance metrics for devices with a single major heat-conducting path is defined.

For all these reasons, the $Z_{\mathrm{thJC}}(t_{\mathrm{p}},D)$ plots reduced to an assumed case surface have very limited practical meaning; they provide solely a rough estimation on the thermal behavior of a device in an actual assembly.

In most power electronics constructions, the portion of the heat-conducting path within the packaged device is of lower thermal resistance than that of the cooling mount. Taking the $Z_{\mathrm{thJC}}(t_{\mathrm{p}},D)$ plots as system model for pulsed excitations would postulate that the package case or module baseplate is at a fixed and known temperature, which practically never occurs; it could be realized with infinite cooling capability on the case surface.

Assuming the other extreme, considering the cooling mount as a mere additional thermal resistance, the actual cooling performance of the assembly at pulsed excitation is severely underestimated. A bulky heat sink can absorb thermal pulses well in the minute range; its contribution to the transitory storage of the heat can be taken into consideration in several steps.

An estimation on the performance of a given heat sink at pulsed excitations can be based on the analytical Eq. (2.51), presented later in Sect. 2.11. As a further step, a

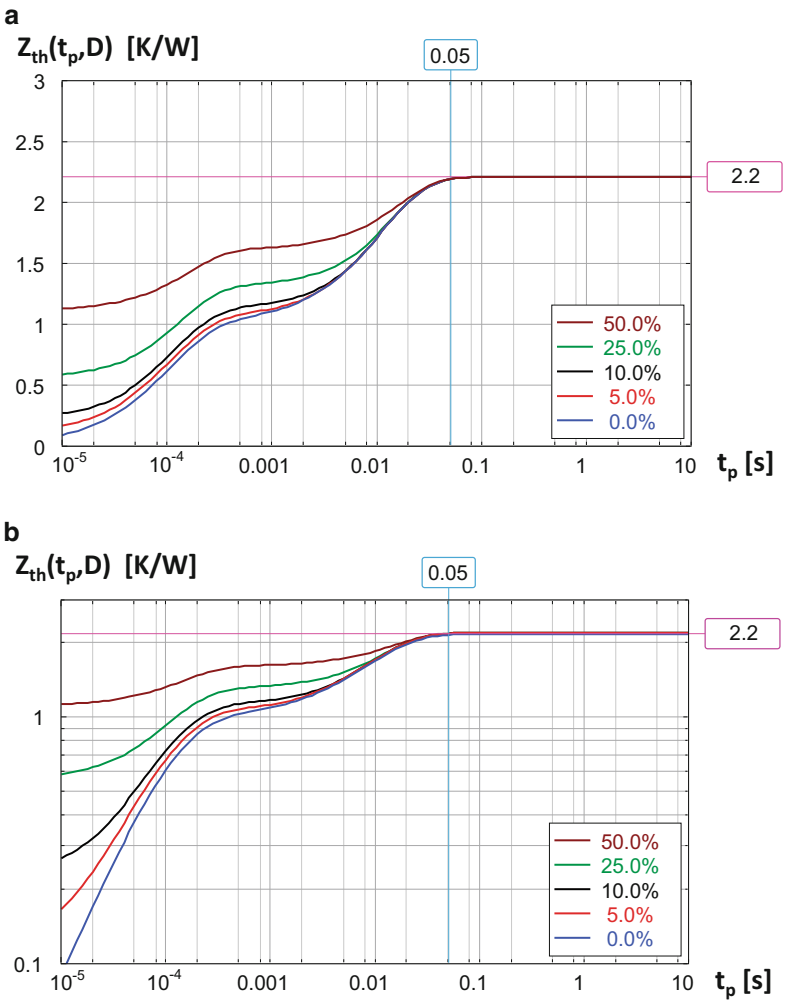

Fig. 2.53 $Z_{thJC}(z,D)$, junction to case pulse thermal resistance of the heat-conducting path scheme of Fig. 2.49, grounded at the Tc node; (**a**) linear, (**b**) logarithmic pulse thermal resistance scale

thermal transient simulation on a detailed model can confirm the suitability of the selected cooling solution.

When the full realized assembly composed of the device with thermal interface and cooling mount is available, it is essential to carry out *single power step measurements* at different powering. These can give an insight into the thermal performance of *all parts* in the assembly, in a wide power, frequency, and duty cycle range, relying on the concept of the pulse thermal resistance.

A more detailed study on pulse thermal resistances is given in [74].

2.9 Relationship Among the Different Representations of Thermal Impedances

To summarize the various representations of the thermal impedance, Fig. 2.54 provides an overview chart that shows also how they are related to each other.

The boxes with lighter blue background are generic to both driving point and transfer impedances.

The time constant spectrum is also generic to both types of the impedances. It has to be emphasized that in case of *transfer impedances*, real, physically meaningful time constant spectra may also include *negative magnitude (R) values*. Practical implementations of the numerical deconvolution algorithm defined by Eq. (2.21) may or may not yield these negative time constant values. This implementation dependence is indicated by the darker blue color of the corresponding box in Fig. 2.54. To be on the safe side, it is better to derive further representations of the self-impedances only from time constant spectra, if the unique properties of the used deconvolution algorithms are not known.

Dark red backgrounds indicate the representations that are defined or used only for driving point thermal impedances. Though the procedures used to obtain the Foster models or the pulsed thermal resistance diagrams do not pose mathematical problems even when the time constant spectra contain negative magnitude values, these representations are mostly used for driving point thermal impedances. Note

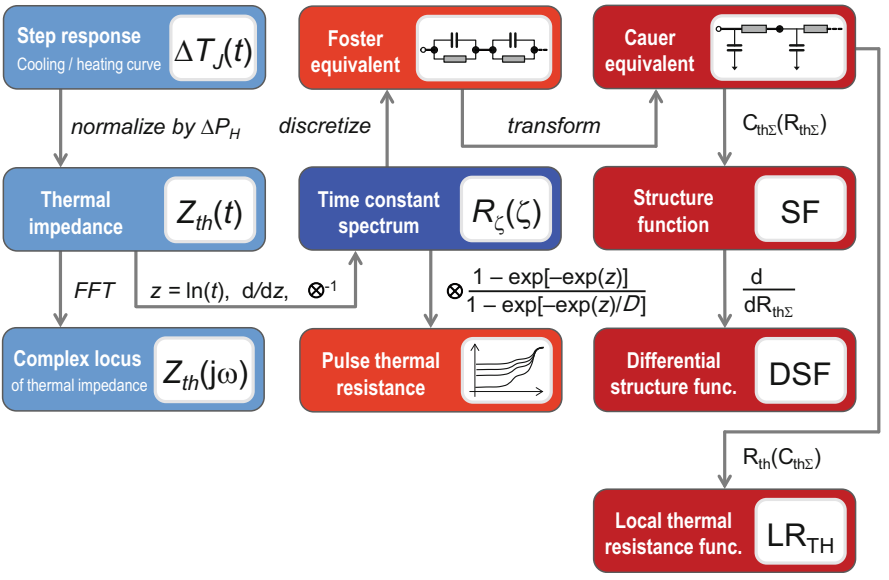

Fig. 2.54 Summary chart of the different representations of the driving point thermal impedances and the transformation paths among them. In light blue boxes, generic representations both for self-impedances (aka driving point impedances) and transfer impedances are shown. In dark red boxes, representations that are defined solely for driving point impedances are presented

that Foster-type network models with negative thermal resistance values can be well handled by circuit simulation algorithms. Note that if the numerical deconvolution algorithm used for the implementation of Eq. (2.21) provides time constant spectra with positive magnitude values only even in case of transfer impedances, the representations shown with any red background in Fig. 2.54 can be numerically calculated. Still, one has to be aware of the fact that parts of the structure functions obtained this way from thermal transfer impedances do not represent physical reality. As the calculation of the complex loci of thermal impedances bypasses the calculation of time constant spectra, a quick look at these loci allows to check if the thermal impedance is a pure driving point impedance or there is some "transfer effect" included therein. If a locus has a section in the quarter of the complex frequency plane with negative real part, that curve represents a transfer impedance, such as the red curve in Fig. 2.47, and the corresponding $Z_{\mathrm{th}}(t)$ function cannot be represented by any version of the structure functions.

2.10 Distributed Heat Transfer on a Surface Towards a Convective Environment

Previously in Sect. 2.5 closed formulae have been constructed for obtaining the thermal resistance and capacitance of finite length beams or tubes. It was assumed in all cases that the heat transfer occurs exclusively at the two ends of the tube; the heat flux enters the tube at one end, at the "driving point" of the equivalent thermal network, and leaves on the other end, at an isothermal surface.

In a convective environment, the cooling of actual heat sinks and cold plates occurs on their whole surface, in a distributed way.

From the viewpoint of thermal transient testing, this distributed heat loss can be an intentional part of the test arrangement terminated by a cooling mount, or it can be an undesirable parasitic effect which distorts the measured thermal quantities.

Such parasitic effects can be parallel heat-conducting paths from the test setup through the surrounding air, or the distortion of the temperature field caused by sensor probes attached to a hot surface. An analytical treatment of the parallel heat-conducting paths and a methodology for a partial reconstruction of the primary path are given in [75–77].

Large chapters of mechanics deal with heat convection, and empirical formulae are listed in the literature which take into consideration the speed of the coolant, possible turbulent effects, and surface roughness and similar. The unintentional parallel cooling in thermal transient testing typically occurs towards the ambient air, and considering a constant h heat transfer coefficient for calculations is mostly satisfactory.

Simulation of detailed system models yields the distribution of temperatures and fluxes for arbitrary geometries and the heat loss at various surfaces towards a convective environment, but the result is valid only for a given geometry and powering.

The analytical formulae of distributed cooling are obtained in a semiempirical way, amalgamating the equations of heat spreading in Sect. 2.1 with empirical correction factors. The treatment of these effects is beyond the scope of this book. However, we illustrate in a short example that the results can be also obtained starting from the discretized RC approach of the previous sections.

Example 2.10: Heat Spreading in Wires and Long Fins in a Convective Environment

In an important practical case, when a temperature sensor is attached to a hot surface, the leads of the electrical connection cause an additional cooling at the measured spot, and in such a way, they distort the measured temperature value. The heat loss occurs on the whole wire surface; the cooling at the very end, at the instrument, can be neglected.

Similarly, plate or pin fins of a heat sink can be considered "long" when most of the heat flux leaves on the surface, before reaching the far edge or tip of the fin.

Assuming infinite length of these structures, simple analytical formulae of their thermal resistance can be obtained again.

Suppose in a Cauer-type ladder network (Fig. 2.55) the series Z_s elements represent the thermal resistance of a section in the fin or wire which impedes the heat propagation towards further similar sections. The parallel Z_p elements correspond to the heat transfer towards the air or other convective environment.

Adding a further Z_s-Z_p pair to an infinite ladder does not change its Z_{in} driving point impedance:

$$Z_{in} = Z_s + \frac{Z_p \cdot Z_{in}}{Z_p + Z_{in}}, \tag{2.46}$$

rearranging and solving for Z_{in} the equation yields

$$Z_{in} = \frac{Z_s + \sqrt{Z_s^2 + 4 \cdot Z_s \cdot Z_p}}{2} \tag{2.47}$$

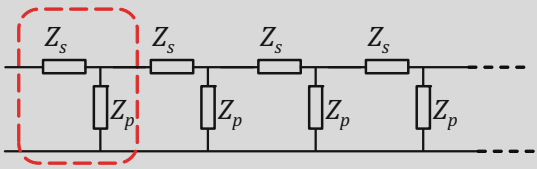

Fig. 2.55 Infinite Cauer ladder

(continued)

Example 2.10 (continued)

Along a dx length of the wire or a pin fin, the infinitesimal thermal resistances can be calculated as

$$Z_s = \frac{1}{\lambda \cdot A_s} dx = \frac{1}{\lambda \cdot r^2 \pi} dx \quad \text{and} \quad Z_p = \frac{1}{h \cdot A_p} = \frac{1}{h \cdot 2r\pi \cdot dx}, \tag{2.48}$$

where r is the radius of the cylindrical element, A_s and A_p are the cross-sectional and lateral surfaces and h is the heat transfer coefficient towards the air.

The physical and geometrical parameters in (2.48) can be cumulated into respective M and N multiplying constants: $Z_s = M \cdot dx$ and $Z_p = N/dx$. For this infinitesimal geometry, $Z_s \cdot Z_p = N \cdot M$, and the result from (2.47) can be written as

$$Z_{in} = \frac{Mdx + \sqrt{(Mdx)^2 + 4NM}}{2}, \tag{2.49}$$

The limit of the driving point impedance becomes $Z_p = \sqrt{N \times M}$ as dx approaches zero:

$$Z_{in} = \sqrt{\frac{1}{\lambda \cdot r^2 \pi} \cdot \frac{1}{h \cdot 2r\pi}} = \sqrt{\frac{1}{\lambda \cdot h \cdot 2r^3 \pi^2}} \tag{2.50}$$

Equation (2.50) is known as "heat loss from infinite fin"; it is a useful formula for calculating the heat removal from a surface when measured by thermocouple or PT100 sensor. The λ thermal conductivity is known from the material composition of the thermocouple or of the connecting wire; similarly, the r radius (or "gauge") is provided by the manufacturers.

The $1/r^{3/2}$ dependence of the thermal resistance in the formula suggests that the thickness of the thermocouple is of eminent importance. Likewise, thermal insulation on wires can diminish the heat removal effect.

A correction algorithm to restore the structure functions of an equivalent heat-spreading scheme, subtracting the parallel heat loss towards the ambient, is presented in [75, 76], and is a realized feature in the software toolset of [54].

Figure 2.56 illustrates the simulated temperature distribution in a K-type (chromel-alumel) thermocouple. The diameter of the welded ball at the tip was supposed to be 0.6 mm; the diameter of both wires is 0.2 mm (gauge 32). The tip touches a cold plate of 30 °C temperature; the upper half of the ball is embedded in thermal grease of 40 W/mK thermal conductivity. The ambient temperature was set

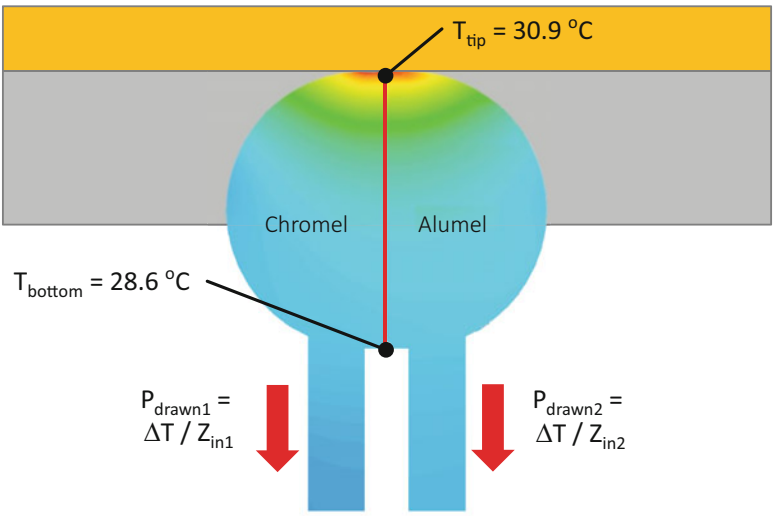

Fig. 2.56 Simulated temperature distribution in a K-type (chromel-alumel) thermocouple. The tip touches a cold plate of 30.9 °C temperature; the upper half of the welded ball is embedded in thermal grease. The temperature at the point where the wires adjoin becomes $T_{\text{bottom}} = 28.6$ °C at $T_A = 20$ °C ambient temperature and $h = 10$ W/m^{2}K heat transfer coefficient towards the ambient. The ΔT temperature difference in the figure is $T_{\text{bottom}} - T_A$

to $T_A = 20$ °C in the simulation, and the heat transfer coefficient towards air was taken as 10 W/m^{2}K.

The thermocouple "reports" the temperature of that point where the two wires adjoin, which is in this case $T_{\text{bottom}} = 28.6$ °C. A larger distance from the surface, typical PTFE insulating coating, and similar other factors may diminish further the accuracy of thermocouple measurements.

2.11 Temporary Heat Storage in the Cooling Mount

In an interesting way, a formula similar to Eq. (2.50) can be used for calculating the thermal impedance of a rod, blade fin, or pin fin at pulsed excitation. The pulsed waveform can be decomposed into an average steady power and a series of power waveforms of frequency f, angular frequency $\omega = 2\pi f$ alternating around the average value.

In this case the Z_s series elements of the Cauer ladder in Fig. 2.55 correspond to the R_{th} thermal resistance of an infinitesimal section of the fin; the parallel component corresponds to the thermal capacitance as $Z_p = 1/\mathbf{s}C$ or $Z_p = 1/j\omega C$. Composing the $Z_s \cdot Z_p$ product, the equivalent thermal impedance of a fin at angular frequency ω will be

$$Z_{\text{in}} = \sqrt{\frac{1}{\lambda A} \cdot \frac{1}{j\omega c_{\text{V}} A}} = \frac{1}{A} \cdot \sqrt{\frac{1}{\lambda j \omega c_{\text{V}}}} \tag{2.51}$$

During system design an estimation on the performance of a selected heat sink can be made based on (2.51).[11] All blade or pin fins contribute to the cooling, as a first assumption their cross-sectional area can be summed up in the formula. As a further step, a thermal transient simulation on a detailed model can confirm that the cooling solution fits the purpose.

A detailed treatment on the effects of external air or liquid cooling on heat sink fins or wires is presented in [79].

2.12 The Limits of the Linearity Assumption

In the previous sections of this chapter, so far the characteristics and behavior of thermal systems have been treated in a fully linear approach. This approach is justified when the investigated system operates in a temperature range where the nonlinearities can be neglected.

The root cause of nonlinearity in the thermal behavior is the temperature dependence of the thermal parameters, namely, of the thermal conductivity and the specific heat in the material layers which are the most exposed to temperature change. Typically, these structural parts include the semiconductor chip, the die attach, and the ceramics or metal base to which the chip is attached.

Farther elements in the heat-conducting path are mostly at lower temperature because of the applied external cooling. The typical cooling solution in electronic systems is convective heat transfer on dedicated cooling surfaces, assured by either air or liquid cooling. The convective cooling mechanisms have inherent nonlinearities, but the detailed discussion of these effects is beyond the scope of this book.[12]

Nevertheless, thermal testing standards dealing with the environmental conditions of the measurements are aware of these, especially in case of natural convection [31] and forced air cooling[13] [34]. For example, in case of a standard natural convection cooling environment, the air temperature has to be measured at the temperature monitoring point of the test chamber in order to assure that the test environment remained stable during a measurement.

[11] This formula is related to the "RMS heat storage for the thermal skin effect," useful for calculating the seasonal temperature change in different depths of the earth. The phase change effect with a periodic thermal signal is also used to measure thermal conductivity, because it is proportional to the thermal diffusivity (3ω method).

[12] These effects are related to the nature of the fluid flow of these media (such as laminar flow turning into turbulent, etc.); discussion of these is the subject of fluid mechanics.

[13] As the CFD-based thermal simulation tools emerged and their use became daily engineering practice in electronics cooling design, the relevance of physical testing of semiconductor device packages under forced air cooling conditions has significantly decreased.

Some special electronic appliances operate at high temperatures, such as vacuum tubes in broadcasting systems or silicon carbide rectifiers in locomotive applications. At these temperatures the investigation of radiation has to be involved, which follows the Stefan-Boltzmann law:

$$\varphi(t) = \varepsilon \cdot \sigma \cdot T(t)^4, \tag{2.52}$$

which tells that the emitted φ heat flux from a surface portion of a hot body is proportional to the fourth power of its T temperature. In (2.52) σ is a physical constant, the Stefan-Boltzmann constant. The material composition and the surface quality also influence the emitted power. This is represented by the ε *emissivity* of the surface; it is $\varepsilon = 1$ for an "absolute black body" and lower for real materials. Shiny metal surfaces have an emissivity below 0.1, while ε of the anodized aluminum is above 0.7. Paints are typically above 0.9, independently from their color, as they have typically the same "color" in the infrared spectral range where emissivity really counts.

Although the radiative heat transfer from the circuit boards or hotter package or heat sink surfaces improves slightly the cooling of regular electronics, these surfaces are typically below 100 °C, where radiation has a minor role. The spectacular blackening of heat sinks serves mainly marketing purposes.

Time-dependent variation of the material properties is also out of the scope of this current discussion, especially since these changes (e.g., thermal conductivity change due to the dry-out of TIM pastes) are slow. Slow in this context means that that the pace of such changes is slower by multiple orders of magnitudes than the lengths of the temperature transients that we aim to measure. This huge difference in the pace of changes allows one to use structure functions to monitor the degenerative (aging) processes in certain structural elements of semiconductor device packages. Such applications of thermal transient testing are discussed in Sect. 7.4.

Accounting for possible nonlinearities is the most relevant for the early parts of driving point thermal impedances since these originate mostly from the temperature dependence of the materials used inside a semiconductor device package [69]. Though these nonlinearities slightly effect the thermal transfer impedances as well (such as the *Chip1-Chip2* transfer impedance for the arrangement presented in Fig. 2.40), they are most affected by the properties of materials outside the package structure – as it will be shown later, e.g., in Chap. 3.

Thermal transient testing may aim at different targets. One purpose can be the determination of the temperature change in time, especially finding the maximum temperature as a crucial factor which influences lifetime. Another target can be checking internal structural details, partial resistances, and assembly integrity.

Therefore, during thermal transient testing, it is worthwhile to use different levels of heating power for different purposes:

1. If the applied temperature sensors and the measurement apparatus are sensitive enough to record tiny signals at proper resolution, lower levels of heating power can be applied that result in a low junction temperature rise. Remaining below

10 °C temperature change, the nonlinearities of the material properties are negligible; thus, using the linear system theory for the postprocessing of the measured thermal transients is well justified.

2. If the purpose of the thermal transient testing is to characterize a device under conditions close to the relevant field applications, the heating power levels should approximately match the levels of the foreseen levels in use. In many cases typical junction temperature elevations are still expected to stay below 100 °C, because of reliability considerations. A case study of exceeding this temperature change limit and stepping into the range where nonlinearities of the material parameters matter will be presented below in Example 2.11. Still, the outcome of the study is that the linear apparatus can manage thermal changes in the 150 °C range.
3. In reliability and accelerated lifetime tests, a common practice is to apply heating power levels beyond the ones usual in field applications. These result in high temperature elevations when nonlinear effects become significant. Such effects may not necessarily hamper the NID method-based postprocessing of the measured transients, but need to be known and properly accounted for, e.g., applying right correction formulae. Even in accelerated tests when the device is continuously stressed with high-power pulses, the variations in the structure functions still can properly reveal when and how fatal device degradations appear and develop. Such applications are treated in Sect. 7.4 of Chap. 7.

Further on in this section, we focus on the second case, on nonlinearities encountered during the measurements in typical operating conditions.

2.12.1 *The Most Common Nonlinearities: Temperature-Dependent Material Parameters*

The major source of the nonlinearity of the thermal systems is the temperature dependence of the λ thermal conductivity and of the c_v volumetric heat capacity of the materials used in the structure. Different materials show rather different temperature dependence, but the alteration of thermal parameters is low in the usual range of the operation of electronics devices (200–400 Kelvin). In [64] a thermal transient technique is presented which yields the temperature dependence of the thermal conductivity of materials used in electronics packaging.

In the usual temperature range of operation, these dependences can be of different nature, but for typical materials, they can be described with an exponential formula:

$$\lambda = \lambda_0 \cdot \exp\left[\alpha_\lambda \cdot (T - T_0)\right], \tag{2.53}$$

where T_o is the reference temperature, λ_o is the thermal conductivity of the material at the reference temperature and α_λ is the coefficient of temperature dependence (CTD) of λ. The α_λ value is nearly equal to the relative change of λ for 1 °C

Table 2.4 Values the λ_o, α_λ, c_{vo}, and α_c for some packaging materials (averages in the 300–400 Kelvin range)

Material	λ_0 W/mK	α_λ 1/K	c_{v0} Ws/m^{3}K	α_c. 1/K
Cu	401	−0.0001	$3.44 \cdot 10^6$	0.0003
Ni	90.7	−0.00012	$3.95 \cdot 10^6$	0.0008
Ag	429	−0.000094	$2.47 \cdot 10^6$	0.00017
Inconel	11.7	0.0014	$3.74 \cdot 10^6$	0.00075
Al_2O_3	36	−0.0031	$3.04 \cdot 10^6$	0.002
Si	148	−0.004	$1.66 \cdot 10^6$	0.001

temperature rise. For small temperature changes, (2.53) can be well approximated with the

$$\lambda = \lambda_o \cdot [1 + \alpha_\lambda \cdot (T - T_o)] \tag{2.54}$$

linear relationship. The values for some common materials of packages are presented in Table 2.4, taken from [64].

In the dynamic behavior, the temperature dependence of the heat capacitances may also play a role. Fortunately, this effect is rather small and often negligible in the 0–150 °C range. For the description of the temperature dependence of the heat capacity, a function similar to (2.53) can be used as

$$c_v = c_{v0} \cdot \exp[\alpha_c (T - T_0)], \tag{2.55}$$

where c_{vo} is the volumetric heat capacity value at the reference temperature and α_c is the coefficient of the temperature dependence. Similarly to the thermal conductivity, the linear approximation holds also here.

Values of c_{vo} and α_c are also presented in Table 2.4; their temperature dependence is shown in Fig. 2.57. As the table shows, these parameters change only slightly in the usual temperature range of operation.

Various authors have examined the significance of nonlinearities in the thermal behavior of electronics packages [64, 69, 80]. They agree that for small temperature changes (<25 °C), the error of the linear approximation is negligible. For temperature changes within the 0 °C, 100 °C range the error is about 2–5%, depending of the materials used in the structure. In case of larger temperature changes, the error can be higher, depending of course again on the α parameters of the different materials in the structure.

As mentioned previously in this section, the temperature dependence of the structural materials of the heat conduction path from the semiconductor junction affects mostly the driving point thermal impedances at early times, as emphasized by D. Schweitzer et al. in [69].

In Eqs. (2.26) and (2.27) in Sect. 2.4.2, the time evolution of the junction temperature transient was expressed with a $\Delta T_J(t) = k_{therm} \times \sqrt{t}$ formula, where both the λ thermal conductivity and the c_v volumetric heat capacity appear in the definition of the k_{therm} coefficient.

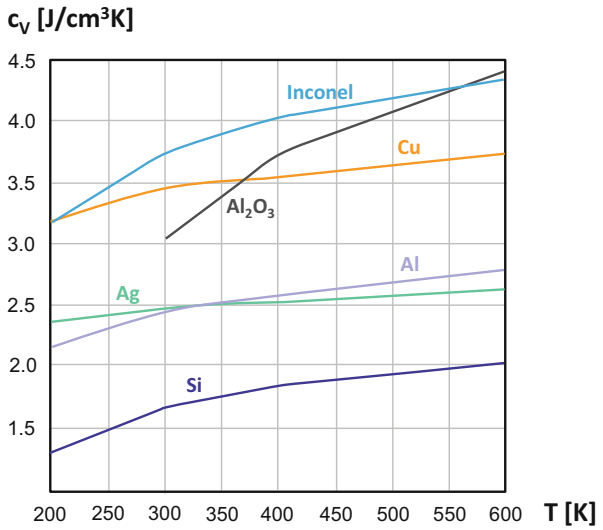

Fig. 2.57 Temperature dependence of the volumetric heat capacity of a few materials frequently used in semiconductor device packaging [3]

As shown in [69], if the $Z_{th}(t) = \Delta T_J(t)/P_H$ driving point thermal impedance is known for an initial T_1 temperature of the chip, then for an elevated T_2 chip temperature, the value of the thermal impedance can be rescaled as follows:

$$Z_{th}(T_2,t) = Z_{th}(T_1,t) \cdot \sqrt{\frac{\lambda_{sem}(T_1) \cdot c_{v-sem}(T_1)}{\lambda_{sem}(T_2) \cdot c_{v-sem}(T_2)}} \tag{2.56}$$

where λ_{sem} and $c_{v\ -\ sem}$ are the temperature-dependent thermal properties of the semiconductor chip.

The above equation and Eqs. (2.26) and (2.27) were derived with the assumption in [69] that the semiconductor chip is thick enough to consider as infinite in size for a prolonged time. Moreover, it was assumed that the heat leaves the junction through the chip, towards the bottom of the structure, and there is no heat flux towards the top. Extending this concept, A. Alexeev et al. recently derived in [77, 78] a new formula for $\Delta T_J(t)$ that is valid for the cases, when the heat generated at the junction flows in both directions, for example, in case of LEDs where some heat also leaves through the lens. Similarly to D. Schweitzer's correction for nonlinearities, also in this case, early transients follow the $\Delta T_J(t) \sim \sqrt{t}$ time dependence; the square-root approach for $Z_{th}(t)$ is maintained.

Based on (2.56), the correction procedure in [69] accounts for these temperature dependencies during thermal simulations. The models of the temperature dependence can be similar to (2.53) and (2.55) or to the simpler (2.54) for the thermal conductivity.

Figure 2.58 presents finite element-based simulation results for a device driven with a short heating pulse of 200 W [69]. The blue curve represents the "real" device behavior with considering the temperature dependence of the thermal conductivity

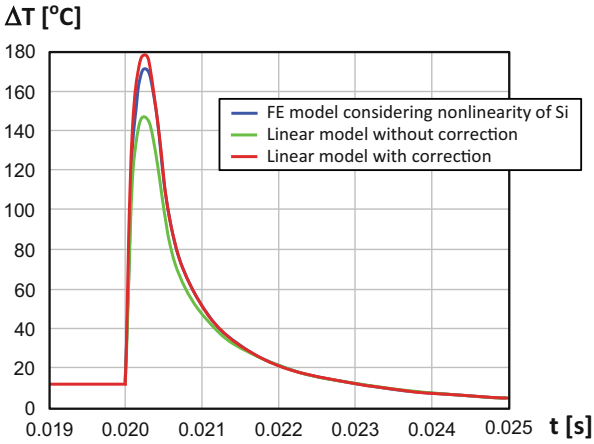

Fig. 2.58 Simulated response of a real multi-chip power device structure (similar to the one in Fig. 2.40) for a short power pulse of 200 W

and volumetric heat capacity of the heated silicon chip in an exact way. The green curve was obtained by a purely linear model-based calculation (thus, with temperature-independent, constant material properties), while the red curve was obtained by calculations where a linear thermal model completed by the correction procedure based on Eq. (2.56) was used [69]. By comparing the green and the blue curves, one can clearly see that at junction temperature elevations above ~150 °C, neglecting nonlinearities results in large (>10%) error.

Note, however, that according to the study reported in [64], if the junction temperature elevations do not exceed ~100 °C, the errors due to neglecting the temperature dependence of the material properties remain below 5%. Thus, the postprocessing of the measured thermal transients using deconvolution or other apparatus based on the linear system theory is justified from an engineering perspective.

Temperature dependence of the effective thermal conductivity of thermal interface layers is another reason why one may observe temperature-related changes in structure functions.

Example 2.11: Nonlinearity of the Thermal Behavior of a Packaged SiC Power Device

A CFD simulation with the tool of [56] was carried out on the detailed model presented in Fig. 2.59. In the arrangement a power MOSFET device in a TO220 package was placed on a cold plate with an inserted alumina sheet of 1 mm thickness.

The chip was modeled as a SiC block of 1.3 mm × 1.3 mm × 0.3 mm size, with a 1 mm × 1 mm × 0.005 mm dissipating junction on its surface. The die attach layer was of 0.025 mm thickness. The simulation was carried out in the 10 °C–90 °C cold plate temperature range, in 20 °C steps.

(continued)

Example 2.11 (continued)

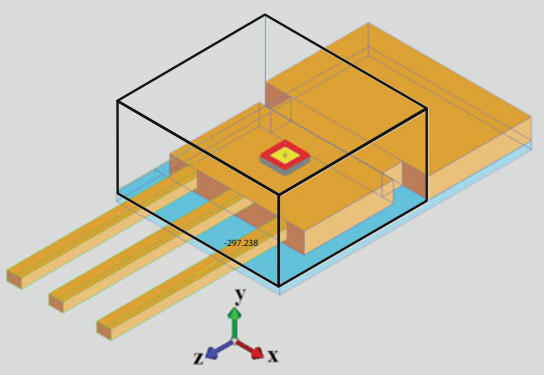

Fig. 2.59 Detailed model of a SiC power transistor for thermal simulation in the SIEMENS SIMCENTER Flotherm tool [56]

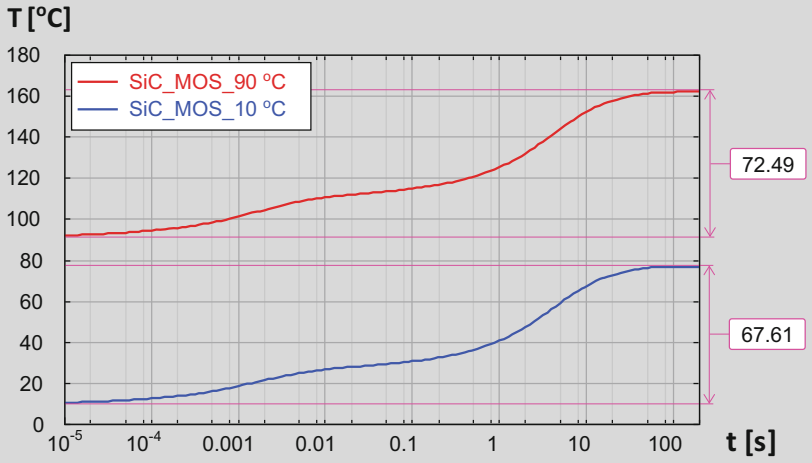

Fig. 2.60 Temperature change of the MOSFET device on cold plate, during heating at applied 10 W power, at 10 °C and 90 °C cold plate temperature

In the simulation partly the material parameters built into the tool, partly data from the literature were used.

The temperature range to be considered was quite broad; at 10 W applied power, the simulated device temperature varied between 10 °C and 160 °C, depending on the cold plate temperature; see Fig. 2.60.

In this range the simulation tool uses a piecewise linear approach for the thermal conductivity of the SiC chip material ($\lambda = 330$ W/mK until 125 °C and

(continued)

Example 2.11 (continued)

$\lambda = 214$ W/mK between 125 °C and 225 °C). We added further data to the material library of the simulation tool for the copper and ceramics layers in the DBC structure.

Copper was characterized with the thermal conductivity value $\lambda = 401$ W/mK along with a negative temperature coefficient of $\alpha = -0.00011$ /K. The die attach material was modeled with the thermal parameters of $\lambda = 40$ W/mK; $\alpha_\lambda = -0.007$ /K. The ceramics layers were represented by Al_2O_3 material, with $\lambda = 36$ W/mK; $\alpha_\lambda = -0.003$ /K values. The c_V volumetric specific heat (volumetric heat capacity) was specified as 3.45 J/cm^{3}K for copper, 1.65 J/cm^{3}K for the die attach, and 3.03 J/cm^{3}K for the ceramics.

The simulated transients were converted to structure functions, as shown in Fig. 2.61. Several regions could be identified in the structure functions based on their thermal capacitance calculated from the volume in the module and the assigned specific heat.

The sections below 1 mJ/K were identified as the SiC chip and until 8 mJ/K as the die attach, denoted as `DA`. Two arrowed lines correspond to the volume of a copper block based on the dimensions of the package base in the model (line `Cu`) and to the third of the volume (line `Cu/3`). The end of the steep section which can be attributed to the high conductivity of copper lies between the Cu and Cu/3 positions. This may refer to a heat spreading in a truncated pyramid between the small chip and the wide package base bottom touching the ceramics, as introduced in Sect. 2.5.

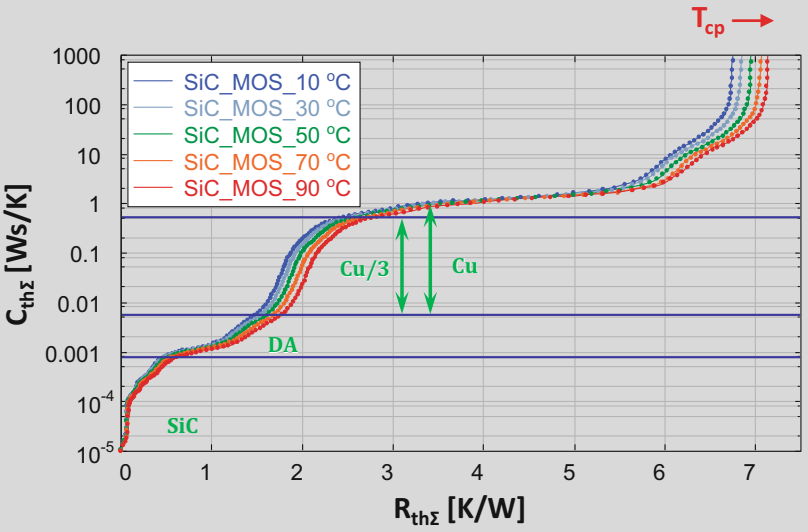

Fig. 2.61 Structure functions of the MOSFET device on cold plate, at applied 10 W power in the 10 °C–90 °C cold plate temperature range

(continued)

Example 2.11 (continued)

It has to be noted that the structure functions gained from real thermal transient measurements are often of lower steepness in the die attach region. This is typically due to die attach voids or delamination. A study on the effect of die attach voids of different coverage is presented in Example 7.5 in Sect. 7.3.

As all constructional materials were supposed to have a negative thermal coefficient, accordingly, growth of the thermal resistance at higher temperature can be observed in all portions in the heat-conducting path. Figure 2.61 proves that the temperature-related difference in the thermal resistance starts building up in the SiC and die attach regions. Fitting the structure functions at the ambient (Fig. 2.62), we can see that the copper and the subsequent layers have a minor share only in the growth of the thermal resistance.

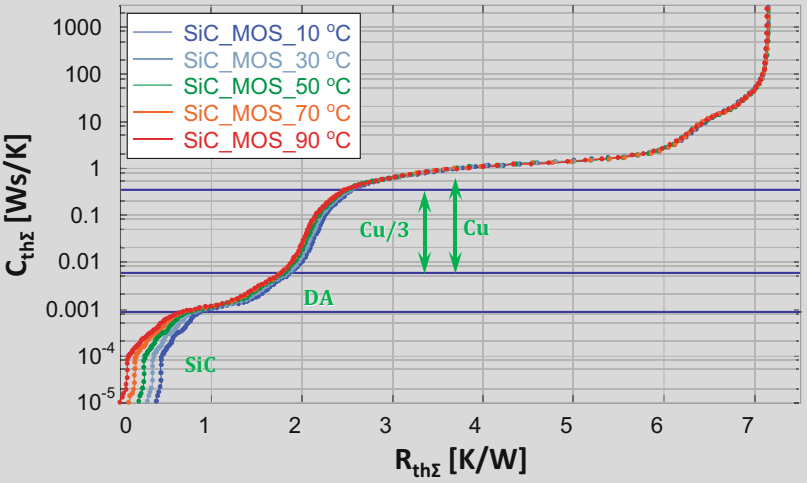

Fig. 2.62 Structure functions of Fig. 2.61 fitted at the thermal resistance of the ambient

Figures 2.63 and 2.64 present the calculated structure functions for the lowest and highest cold plate temperatures, when both heating and cooling transients were simulated. The orange and red curves denote the results of *heating*, for 10 °C and 90 °C cold plate temperatures, respectively. The results calculated from cooling transients are represented as blue and gray plots.

(continued)

Example 2.11 (continued)

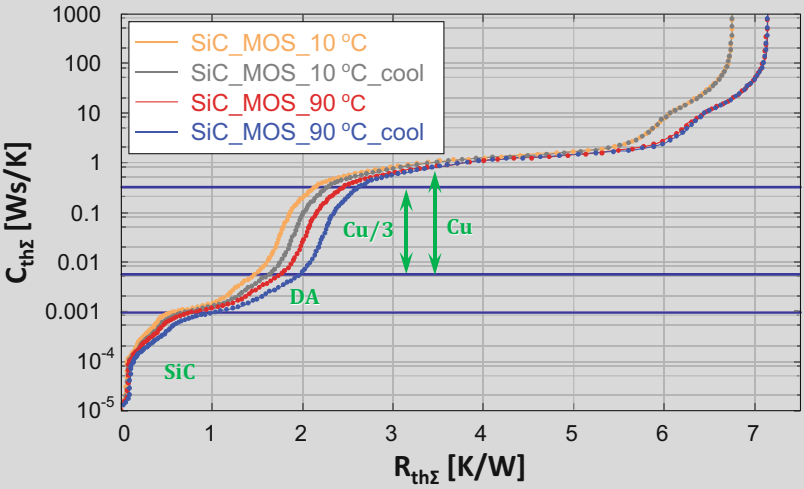

Fig. 2.63 Structure functions derived from simulated heating transients (red and brown curves) and cooling transients (blue and green curves), at cold plate temperature 10 °C and 90 °C

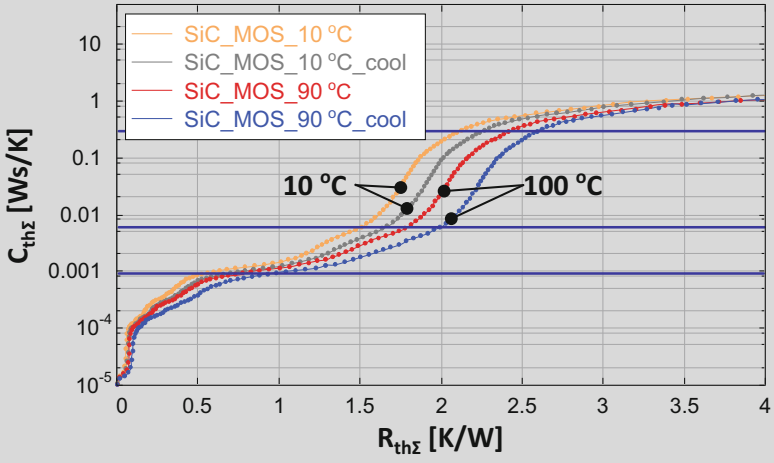

Fig. 2.64 Excerpt of Fig. 2.63. Change of the thermal conductivity in the SiC, die attach, and copper region can be observed

In the previous example, the structure functions calculated from heating and cooling measurements match well along the heat-conducting path. This suggests that for this device, cooling or heating transients yield nearly the same data for structural analysis, despite the assumed nonlinearities. Accordingly, using the linear system theory for data processing is well justified.

It can be observed in the plots that as all material layers were expected to have a thermal conductivity of negative thermal coefficient, all regions in all structure functions shift towards higher thermal resistance values at higher temperatures. This indicates a possible instability, which may result in thermal runaway. In a thermal assembly design, increased cooling capability of the cooling mounts at higher temperatures has to be ensured in order to guarantee thermal stability.

This highlights the fact that in case of identifying the so-called standard thermal metrics of packages, the test environment (applying a layer of thermal interface material in this example) and the test conditions[14] (the cold plate temperature here) have a significant effect on the overall junction to ambient thermal resistance we measure. This also suggests that one needs to apply well-defined and easily repeatable and reproducible procedures to separate the structure function portions corresponding to the package under test and corresponding to the test environment.

2.12.2 Measurement Artifacts Appearing as Nonlinearities

In our discussions so far, we assumed that ideal measurement apparatus was used, i.e., all the captured signals represent the true temperature transients. In practice, this is not always the case. It strongly depends on the way how the temperature is measured by the thermal transient test equipment. Details of usual realizations of such equipment are discussed in Chap. 5, followed by Chap. 6, providing descriptions of measurement basics of different types of electronic components. The discussion here is restricted solely to the possible artifacts caused by the actual temperature measurement method. These – if not known – are wrongly appearing as the nonlinear behavior of the thermal system realized by the device package.

In daily practice direct measurement of temperature is replaced by indirect methods, matched to the actual temperature range of interest. Early thermometers used to measure temperatures in everyday human environment were based on the physical effect of thermal expansion, converting it to a length scale. Modern, electrical thermometers convert the temperature to an electrical signal, e.g., to voltage. The accuracy of the practical temperature measurement depends on how accurately a known temperature change is calibrated against the change of the electrical signal of the electrical thermometer.

[14] The so-called standard thermal metrics and issues of the related thermal test conditions will be discussed in Chap. 3 in detail.

In thermal testing of electronics, the temperature change is converted to electric signal by active semiconductor devices used as temperature sensors using one of their temperature-sensitive parameters (TSP), or by other dedicated electrical temperature sensors attached to accessible surfaces of the measured system. A few important ones are:

- *Diode sensors*
  We usually say that the T_J junction temperature dependence of the V_F forward voltage of a diode driven by a constant forward current is linear, but this is also just an approximation, which is true only for small temperature excursions (<50 °C). In Chap. 4 the actual sensitivity is derived from semiconductor physics, and the resulting equation shows that, e.g., in a range of 200 °C, this dependence is far from linear. The nature of nonlinearity of the $V_F(T_J)$ relationship is more pronounced, e.g., for III–V compound semiconductors than for silicon.
- *Resistor sensors*
  In resistor-based temperature sensors, the temperature dependence of the electrical resistance is utilized. Such thermometers also frequently show nonlinearities. For example, the resistance of metal sheets is of exponential temperature dependence. For example, a PT100 platinum sensor has 100 Ω resistance at 0 °C, and it grows by 385 ppm/K with the temperature. In a small temperature range, this means that applying 10 mA electric current to sense the temperature, the obtained voltage will be 1 V, with the sensor exhibiting 3.85 mV/K sensitivity, but in a broader range, the sensitivity will change, corresponding to the sensor's exponential temperature dependence.

In both cases, in simple measurements, a single sensitivity parameter is used that is a good approximation of the sensors' real characteristics only for a relatively small temperature range (e.g., $\Delta T < 50$ ° C). If the temperature elevations are beyond the validity of linear approximation of the temperature sensors' characteristics, but still a single sensitivity parameter is used for the temperature-voltage conversion, the measured $\Delta T_J(t)$ transients and the corresponding $Z_{th}(t)$ thermal impedances will be distorted by the measurement error, and the measured thermal system would appear as if it was nonlinear. Such a nonlinearity of a thermal system is obviously a *measurement artifact*.

If during thermal characterization of a semiconductor device larger temperature ranges are involved (e.g., $\Delta T > 50$ ° C), a careful calibration of the sensors and the exact calculation with the temperature sensors' actual characteristics is a must to avoid the above measurement artifacts. For example, the software of [54] supports such careful sensor calibrations and performs polynomial or exponential fitting on the set of measured points; the resulting calibration files can be used by the data postprocessing software to yield $Z_{th}(t)$ thermal impedances free of the abovementioned artifact.

Example 2.12: Forward Voltage: Temperature Mapping of a Power LED Device

Figure 2.65 shows a calibration curve of the voltage-to-temperature mapping (calibration curve) of a Royal Blue Cree XP-E2 medium power LED device driven by 10 mA constant forward current. The temperature range covered by the presented set of calibration data is 160 °C. In the diagram we also present results of two different linear approximations for the low and high temperature ranges.

If the sensitivity parameter derived by linear regression for the low temperature range is used, e.g., for a junction temperature above 50 °C (resulting in $\Delta T_J > 60$ ° C) from the reference, the low temperature region), significant errors will develop. To avoid artifacts in the postprocessing of the measured $\Delta T_J(t)$ transients of the LEDs, a quadratic approximation of the $V_F(T_J)$ relationship, as shown in the figure, is already sufficient [81, 82].

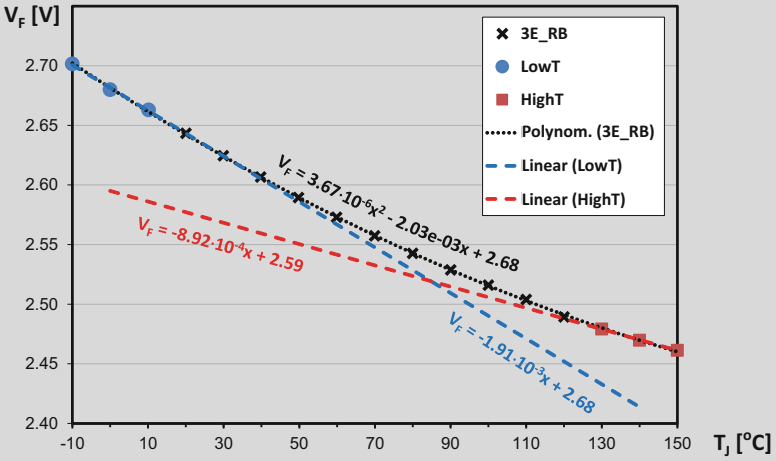

Fig. 2.65 Voltage-to-temperature calibration curve of a Royal Blue Cree XP-E2 medium power LED device at 10 mA sensor current. (Based on Refs. [81, 82])

2.12.3 *Limits of the Validity of the Linear Approach in Actual Measurement Results*

So far this chapter has elaborated the theoretical aspects of heating and cooling of a thermal system. In further chapters the implementation and evaluation of actual measurements will be discussed.

Nowadays all thermal measurements are based on recording temperature-related electric signals of the chips in an appliance or of dedicated temperature sensors. The recorded signal has to be examined in all time intervals for the following criteria:

- Is the electric transient in a time interval related to a thermal change of some parameters or is it caused by other electric effects?
- If it is of thermally induced nature, does the related thermal change occur within the investigated thermal subsystem such as a tested device or module, or in a broader environment out of the device?
- If it is within the investigated thermal subsystem, is its behavior linear in the thermal domain?

In subsequent chapters, especially in Chap. 5 it will be expounded how high power is generated in different device categories such as transistors, diodes, or integrated circuits and why is some power maintained on these devices also during cooling measurements. Without going into details, at this point it can be stated that a common way of powering is to force a higher electric current through the device, and a general way of measuring the temperature of an active device is to record some electric device parameter at a low current bias, called the measurement current. The applied power can be always determined from known currents and measured voltages in the electric system.

Multiplying a measured, often called "raw" electric transient by an appropriate scale factor, a sort of "quasi thermal transient" can be gained. The calculation of the power for several device categories and the definition of an appropriate "voltage change to temperature change" conversion factor are presented in Chaps. 5, 6, and 7.

Example 2.13: Measurement of a Power MOSFET Device at Different Powering and Boundary Conditions

Forced current of various levels such as 1 A, 1.5 A, and 2 A was applied on a packaged power MOSFET (IRF540) for 10 seconds. The resulting power on the device was 0.74 W, 1.14 W, and 1.6 W at the three current levels. After revoking the heating power, the change of the voltage on the device was recorded at different measurement currents (100 μA, 1 mA, 10 mA). The measurements were repeated with the device placed on dry and wet cold plate.

Figure 2.66 presents the "quasi temperature" curves recorded at a few selected combinations of heating current, measurement current, and boundary. Three curves in the figure, denoted with a key starting as `wet_D_`, were measured on the wet plate, and the other three on the dry one. The voltage-to-temperature conversion factor was determined for all measurement currents in a calibration process as defined above in Example 2.12 and later in Sect. 5.6.2.

As introduced before, the thermal impedance is defined as the temperature change in time divided by the applied power. The Z_{th} curves of a linear thermal system are identical at different power levels and starting temperature.

(continued)

Example 2.13 (continued)

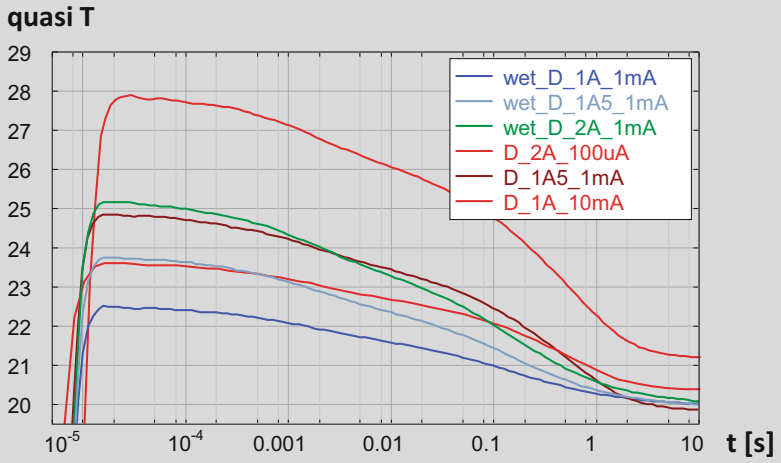

Fig. 2.66 "Quasi temperature" curves calculated from thermal transients of a power MOSFET at various powering and at two boundary conditions. At applied heating current of 1 A, 1.5 A, and 2 A, the power which developed on the device was 0.74 W, 1.14 W, and 1.6 W, respectively

A sort of "quasi thermal impedance" can be produced dividing the transient change of the "quasi temperature" by the applied power. In Fig. 2.67 the curves of this "quasi thermal impedance" are shown, gained from Fig. 2.66 with dividing by the corresponding power. It can be observed that between 30 µs and 100 ms seconds, all curves coincide, regardless of the boundary condition; then, until 10 seconds those curves coincide which were measured at the same "wet" or "dry" boundary. This indicates that the variation in the recorded "quasi thermal impedance" was of purely thermal root cause after 30 µs; it was proportional to the applied power only. The difference in the thermal quality of the "wet" or "dry" cold plate causes the divergence of the curves after 100 ms; this can be used for identifying the role of the device as opposed to the test environment in the thermally induced temperature change.

Figure 2.68 shows the early section of the "quasi Z_{th}" curves enlarged. All presented "wet" curves in the previous figure were taken at 1 mA measurement current, as it was impossible to find any difference between them at the present resolution of the charts; for better clarity only one of them is shown now.

(continued)

Example 2.13 (continued)

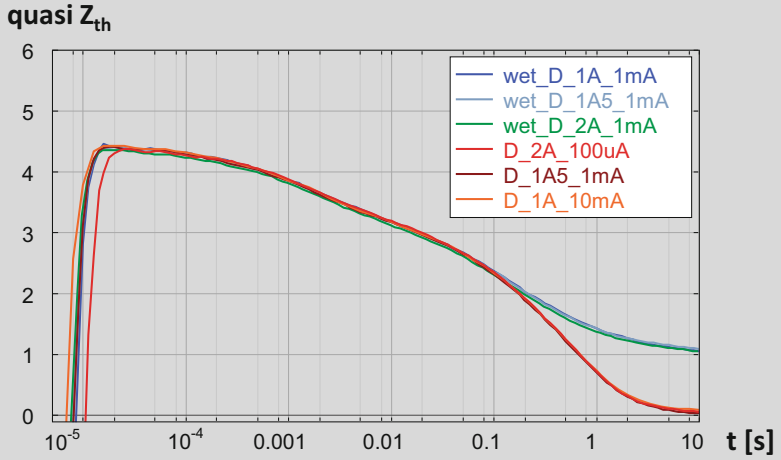

Fig. 2.67 "Quasi Z_{th}" curves calculated from thermal transients of a power MOSFET at various powering and at two boundary conditions. Pure thermal nature of the variation can be seen between 30 µs and 10 seconds where all curves measured at the same boundary condition coincide

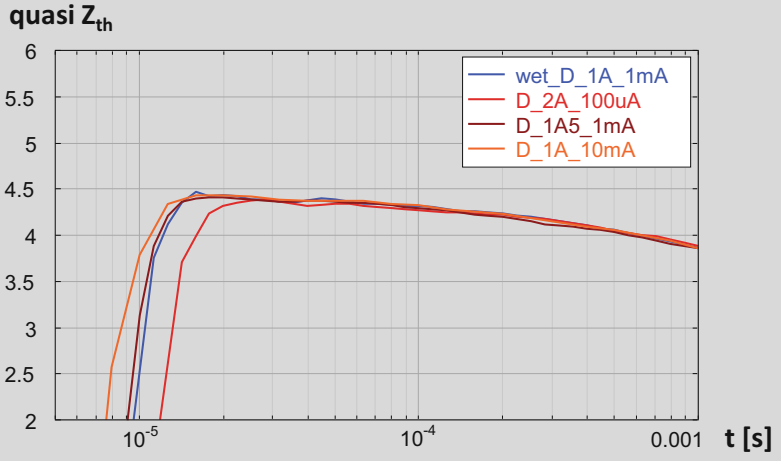

Fig. 2.68 Excerpt of Fig. 2.67. The voltage change on the device is of purely thermally induced nature when the curves belonging to different heating and measurement currents coincide

The range in which the transient change is of the thermally induced nature can be determined with the comparison of transient curves at different powering and measurement currents, as shown in Example 2.13. Omitting the part of nonthermal root cause, typically described as "electric transient"; the valid, true Z_{th} thermal

impedance curves can be gained from the "quasi" Z_{th} curves. Based on Eq. (2.26) even the restoration of the thermal change covered by the "electric transient" becomes possible, as presented in Chap. 6, Sect. 6.1.4.

The distinction between thermal and nonthermal changes in the "quasi" thermal impedance curves is especially complex in the testing of devices based on GaN material (Chap. 6, Sect. 6.9.2). Depending on the measurement technique, these devices can produce electric transients in the many millisecond range which can be mistakenly attributed to thermal effects. A study on the separation of transient changes of different nature and correction procedures is given in [83].

Effect of the Voltage Drop on Wiring

A sort of "procedural" nonlinearity can be observed in measured Z_{th} curves of large power modules at high heating current. As defined in several standards, linear thermal descriptors of a device such as thermal impedance curves and structure functions are composed from recorded transient curves normalized by the applied power. However, the power value is calculated from the forced heating current multiplied by the voltage, measured on the external pins of the module. The wiring within the module is farther from the active devices and is not heated in the same manner as those are, still, the power fraction due to the voltage drop on the wires is added to the power used for normalizing. This increase in the calculated power which actually does not appear in the heating of the active devices causes a characteristic shrinking of the Z_{th} curves and the structure function at larger heating currents. Although this effect seems to be an artifact at the first glance, it is rather a feature caused by the definition of the thermal measurement standards. The phenomenon is discussed in this context, because it affects the structure functions in a similar way as the nonlinear effects caused by the temperature-dependent material parameters.

Effects of Additional Energy Transport

A similar "procedural" nonlinearity can be attributed to devices with multiple energy transport, like LEDs in which the applied electric power is partly dissipated as heat, partly emitted as light. The emitted optical power does not contribute to the temperature elevation of the device. The LED efficiency that is the share of the emitted optical power compared to the electric power fed into the device strongly depends on the current and temperature. Accordingly, in case the thermal transient measurement aims at the structural analysis of the device, the optical fraction in the power has to be measured and subtracted. Similarly, when an external cooling mount is added to the device, only this "optically corrected," reduced heating power is to be taken into consideration.

This concept is illustrated in Fig. 2.69 showing the structure functions of a Cree MCE LED module on the temperature-regulated device holder plate of an optical integrating sphere. The voltage-to-temperature mapping of the LED was similar to the one shown in Fig. 2.65; it was an obviously nonlinear one.

Still, converting the temperature-induced forward voltage change to temperature by the exact mapping and subtracting the optical power, all structure functions taken at different cold plate temperature fit perfectly in region (2.1) of the figure, belonging to the LED and its metal core heat distributor board. The thermal system of the LED

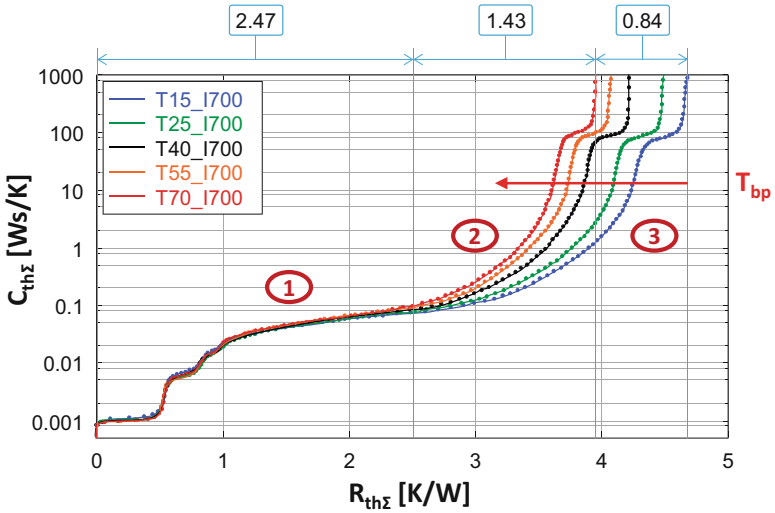

Fig. 2.69 Optically corrected structure functions of a 10 W LED module on the temperature-regulated device holder plate of an optical integrating sphere. Plate temperature T_J is elevated from 15 °C to 70 °C

module can be considered linear. The shrinking of the structure functions in regions (2.2) and (2.3) corresponds to the thinning of the thermal interface between the module and holder plate, and the higher heat sinking capability of the temperature-controlled holder at higher temperature.

Details of the optical correction procedure are presented in Chap. 6, Sect. 6.10.

Summary of Nonlinear Effects

A complete steady-state to steady-state $\Delta T_J(t)$ junction temperature transient comprises all information about the junction to ambient heat transfer, including the heat transfer processes in the test environment, also beyond the conduction path(s) of the package or module under test. Such mechanisms are, e.g., convective cooling from the thermal test board holding the package or, in extreme cases, radiative heat transfer from hot package surfaces. These mechanisms exhibit temperature dependence: the natural convection from hotter surfaces is more intensive; the radiative heat transfer exhibits very strong temperature dependence according to the Stefan-Boltzmann law. In both cases the nonlinearities appear at the largest time constants of the system that can be clearly identified as changes at the very ends of the structure functions and, thus, can be well separated from the parts that are characteristic of the package under test [156].

Treating the thermal effects which are characteristic to the test environment is beyond the scope of this book. It has to be mentioned that providing a correction formula to account for these effects in a linear model-based approach is not straightforward.