SIEMENS

Thermal Behavior of Power Modules in PWM-Inverter

Abstract

In the following, the thermal behavior of power modules in PWM-Inverters will be examined. The starting point is an analysis of single-chip and hybrid structures. A parameter for the characterization of the thermal behavior is the dynamic response in time of the thermal impedance between power semiconductor and device header (Z_{th} curves). For a given substrate configuration, the absolute values of the dynamic response change as a function of chip area, but the time constants do not. This behavior makes it possible to scale the Z_{th} curves, and thus to establish the basis for generating a thermal equivalent circuit for the simulation of chip temperatures in PWM-Inverters. A new aspect of the thermal equivalent circuit is the inclusion of the effects of mutual thermal coupling between individual chips. These effects can be described by means of coupling impedances in combination with the average chip dissipation. Simulations have shown that this mutual thermal coupling can be treated like a static increase of the chip temperature.

At a given dissipation balance, the maximum chip temperature in PWM applications decreases with increasing output frequency, because of the capacitive behavior of the structure. For optimum design of the cooling system it is necessary, therefore, to determine the chip temperatures with good accuracy. By using the method presented here, quick and accurate prediction of the temperatures can be achieved.

Keywords

Power modules, Thermal simulation, Power losses, PWM-Inverter

1 Thermal Behavior of Power Modules

1.1 Structure of Power Modules

The main criterion in the design of packages for power modules is to achieve good thermal conductivity, and a dielectric strength in accordance with the relevant standards, for the insulation between the power semiconductor and the device header. Both parameters are primarily determined by the thermal and electrical characteristics of the substrate on which the semiconductors are mounted.

A widely used mounting technique is **D**irect **C**opper **B**onding (DCB mounting). In this method, a thin film of copper is applied to the top and bottom surface of an electrically insulating substrate with good thermal conductivity by means of a eutectic melt. Chips and base plate are attached to this substrate by soldering.

The power modules have different thermal behavior depending on the substrate material used. Materials such as beryllium oxide and aluminium nitride have particularly good thermal conductivity. However, beryllium oxide is now hardly made use of because of its toxicity, and aluminium nitrides are very costly. As a result, aluminium oxide has established itself during the last few years as the material used by most manufacturers. **Figure 1** shows the structure of a Siemens power module and the thermal parameters of the individual module layers. The values in the brackets are the thermal conductivities. When the power module is mounted on

the heatsink, convex or concave distortion of header and heatsink can produce air gaps which cause deterioration of the thermal transfer impedance between module base and heatsink. During mounting, these air gaps are filled with a paste formed by a suspension of metal oxide powder in oil (heat-conducting paste). With the recommended torque applied to the screw connection and using a thin coating of paste, the Siemens 3-phase bridge module ECONOPACK 2 achieves a layer thickness for the heat-conducting paste of 25 - 40 µm.

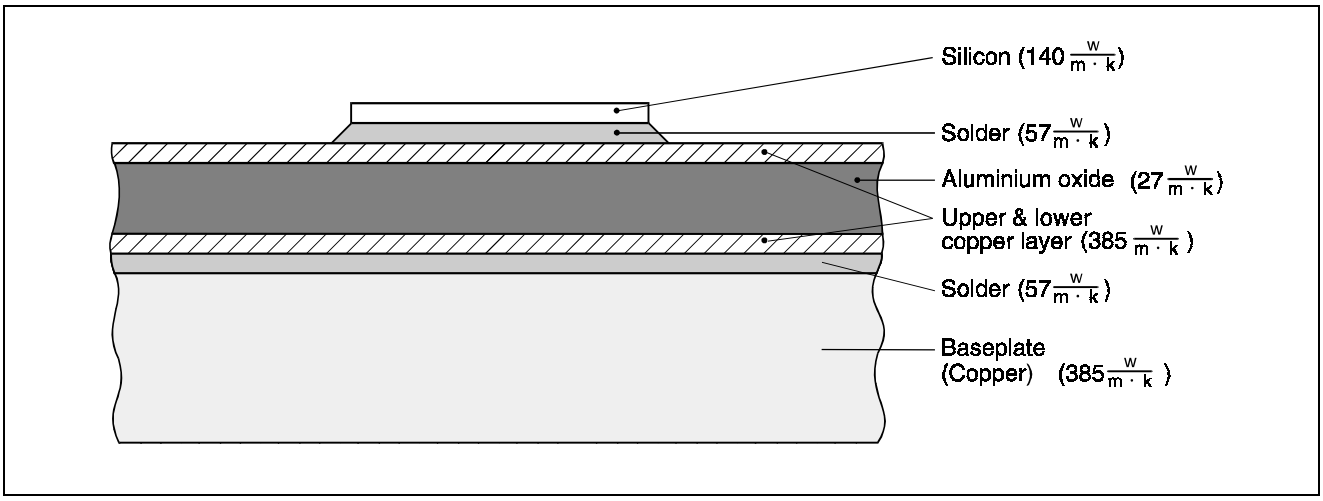

Figure 1
Structure of a Siemens Power Module

1.2 Behavior of Single-Chip Structures

The single-chip structures analyzed in the following are based on a sequence of layers as shown in **Figure 1**. One chip on the substrate functions as the heat source.

The given geometric configuration is divided up into a large number of very small elements (finite element method). For each of these elements, the heat conduction equation given in equation 1 is solved using the ANSYS simulation program.

$$c \times \rho \times \frac{\partial T}{\partial t} = \text{div}\,(\lambda \times \text{grad}\, T) + p_v \qquad \textbf{equation 1}$$

c - specific thermal capacity [Ws/gK]

ρ - density [g/m^3]

T - temperature ($T = f$ (x, y, z.t)) [K]

λ - thermal conductivity [W/mK]

p_v - thermal source density per unit of volume [W/m^3]

The thermal behavior of power modules can be characterized by the thermal impedance between the chips and the header. In analogy to an electrical circuit, this thermal impedance can be calculated from the ratio of the temperature difference between chip and base plate to the dissipation impressed on the chip. **Figure 2** shows the dynamic response of the

thermal impedance (Z_{th} curve) for various chip areas. This dynamic response is obtained as the response to a rectangular dissipation pulse applied between chip and header.

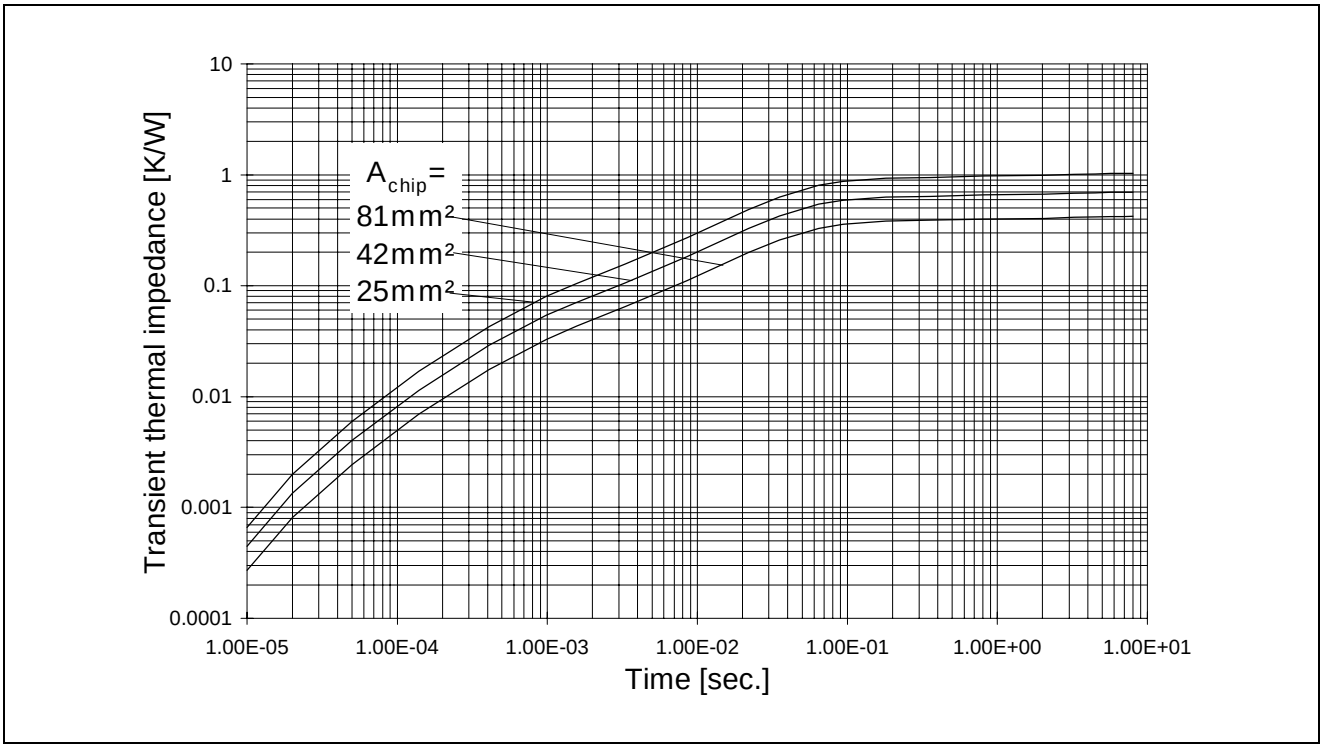

Figure 2
Dynamic Thermal Impedance

Because of the thermal capacities in the structure, the thermal impedance is at first very low and then increases with time. For $t \rightarrow \infty$ these curves approximate the static thermal resistance $R_{th/j\text{-}c}$. Each individual Z_{th} curve can be reduced to a sum of exponential functions and can be calculated by means of equation 2. For different chip areas, the curves can be parallel shifted with respect to each other. The absolute values of the exponential functions vary with chip area, but not the time constants. The dynamic thermal impedances for different chip areas can thus be calculated from equation 3.

$$Z_{th/j-c}(t) = \sum_{i=1}^{n} R_{th/j-c,i} \times (1 - e^{-\frac{t}{\tau_i}}) \qquad \textbf{equation 2}$$

$$\frac{Z_{th1/j-c}(t)}{Z_{th2/j-c}(t)} = \left[\frac{A_{Chip2}}{A_{Chip1}}\right]^{0.76} \qquad \textbf{equation 3}$$

The exponent 0.76 is a measure for the variation of the thermal impedance as a function of chip area for the given configuration. In the same way, this equation can be used for calculating the static thermal resistances ($t \rightarrow \infty$).

The heat-conducting paste used for filling the airgaps between module base and heat sink is a very bad thermal conductor compared to the various layers of the module structure ($\lambda \approx 0.8$ W/mK). Consequently, it has a significant influence on the thermal behavior of the

structure and must be taken into account in the design of the cooling system. **Figure 3** shows the ratio of the thermal resistance due to the heat-conducting paste to the total resistance junction-to-heatsink, in percent. The rise in the relative contribution of the heat-conducting paste with increasing chip area is due to heat build-up (hotspot effect).

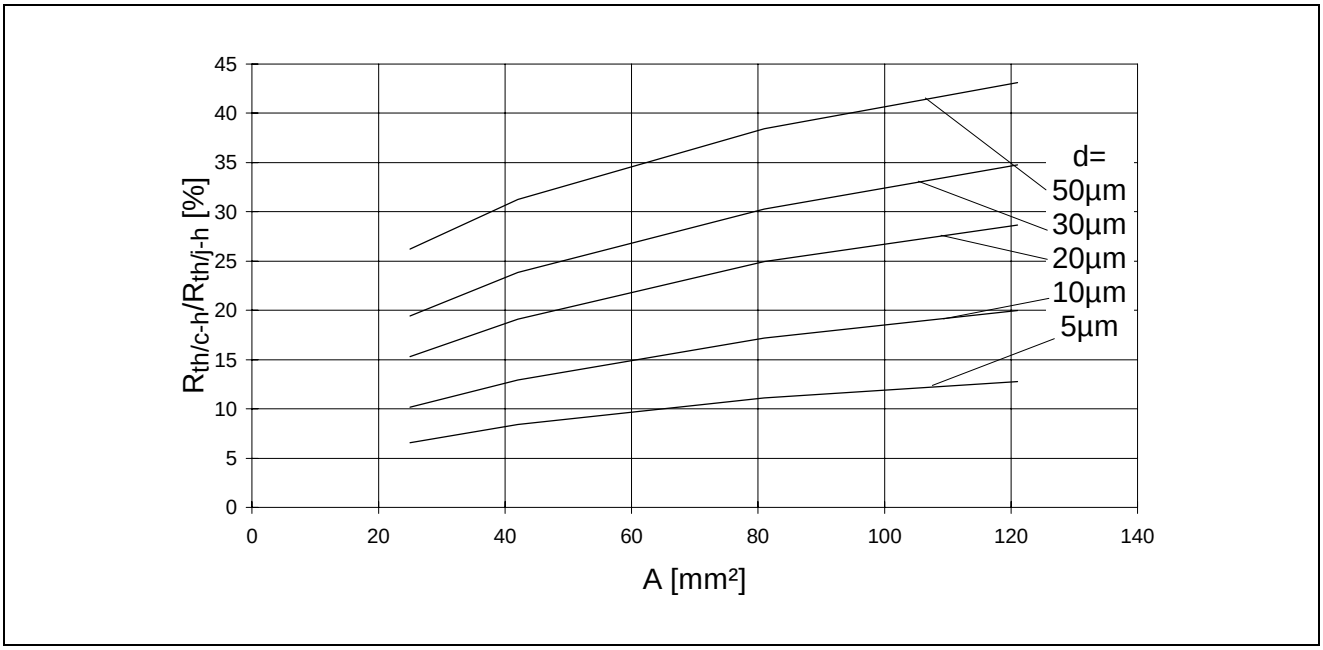

Figure 3
Contribution of Heat-conducting Paste $R_{th/c-h}$ to Overall Thermal Resistance $R_{th/j-h}$, in Percent.

If a dissipation is impressed into the chip, the thermal current from the heat source spreads laterally as well as vertically in the direction to the header. The structure can be viewed as a network formed from a multiplicity of lateral and vertical thermal impedance elements. The thermal field spreads out (heat spreading effect) in a manner depending on the ratio of lateral to vertical network components (Zwanziger /1/). The spreading angle of the thermal current in the module substrate, and thereby the thermal impedance, depends strongly on the chip area. A larger chip area can be regarded as made up of several smaller sub-chip areas. In the inner parts of the chip, mutual thermal coupling between these imaginary sub-chip areas occurs (hotspots). Only in the peripheral areas of the chip is unimpeded lateral flow possible for the thermal current.

As the chip area increases, the ratio perimeter/chip area is reduced. The spreading angle of the lateral thermal current becomes smaller. This increases the relative contribution of the heat-conducting paste to the total thermal transfer impedance.

The hotspot effect is also the cause for the nonlinear change of thermal impedance with chip area (equation 3). The impedance decreases logarithmically with increasing chip area.

A breaking down of large chips into a parallel connection of smaller chips is thermally more favourable. Assuming the same temperature differences between chip and base plate, better utilization of the chips (higher dissipation) is possible.

1.3 Mutual Thermal Coupling in Hybrid Structures

With dissipation occurring in two or more chips arranged in hybrid configuration on the same substrate, the lateral heat spreading effect can cause thermal interaction between the chips. In the regions where the thermal currents are mutually coupled, hotspots occur. The generated heat can not be conducted away properly. This results in an increase in chip temperature. In the following, these mutual couplings are analyzed within a hybrid configuration.

The configuration used for this is that of the substrate part of a Siemens 3-phase bridge module shown in **Figure 4**. In **Figure 5**, the thermal impedance to ambient from along the upper copper surface $R_{th/cu\text{-}a}$ is shown. The module base is terminated with a 30-µm-layer of heat-conducting paste in this case. The temperature underneath this paste is assumed to be constant. The dissipation is impressed in IGBT 2. In accordance with **Figure 4**, the origin for coordinate y is in the center of the chip.

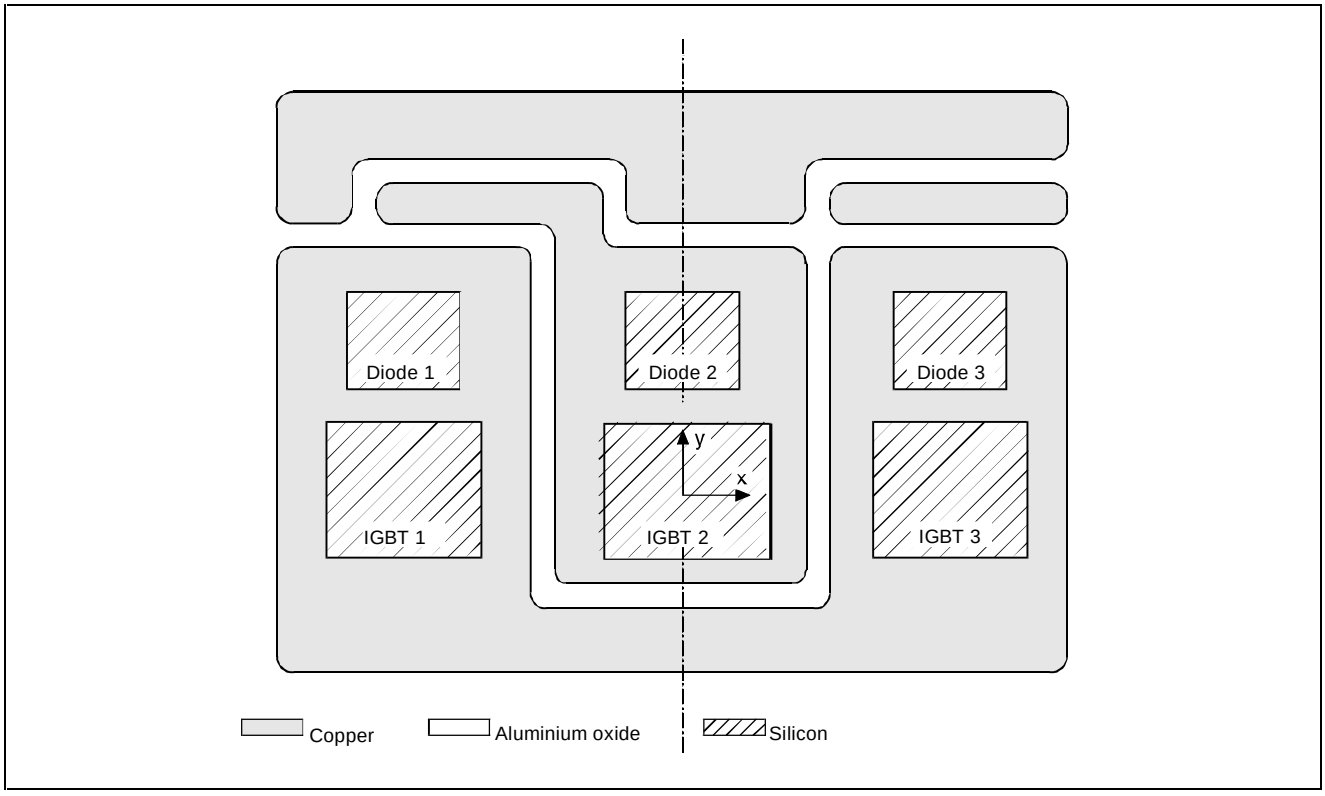

Figure 4
Substrate Part of a Siemens 3-phase Bridge Module (ECONOPACK 2)

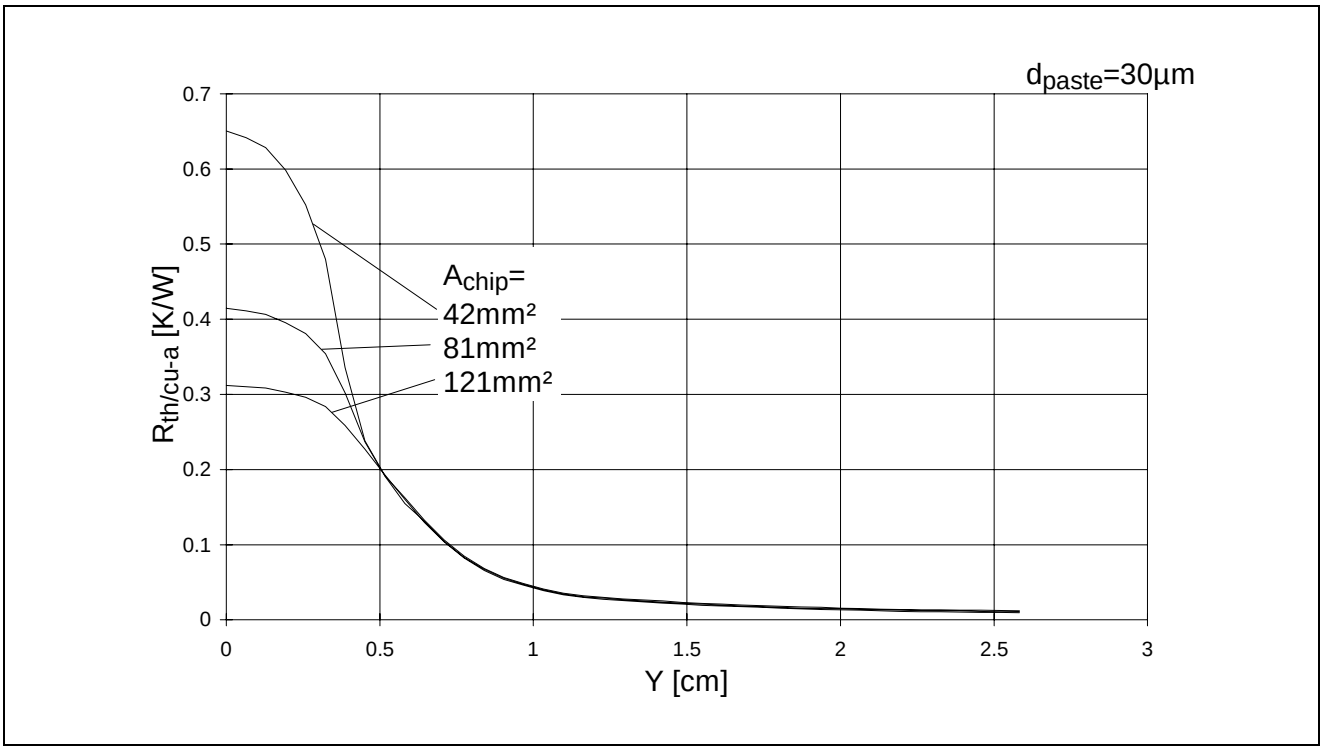

Figure 5
Static Thermal Impedance Along Upper Copper Surface

With the aid of these curves (coupling impedances) and the mean dissipation $\overline{P}_{L,}$ the mean chip temperature increase $\overline{T}_{coup}$ due to mutual coupling in hybrid configurations can be calculated. For example, the mean mutual coupling temperature of IGBT 2 resulting from a power impressed in diode 2 is obtained in accordance with equation 4.

$$\overline{T}_{coup/Diode\ 2 \rightarrow IGBT\ 2} = R_{th/j-a}(y_{Diode\ 2}, A_{Chip/Diode}) \times \overline{P}_{L/Diode\ 2} \qquad \textbf{equation 4}$$

If other chips contribute to the overall dissipation, the overall coupling temperature of IGBT 2 is the aggregate of the partial couplings of the individual elements affecting IGBT 2.

For dimensioning the heatsink, it is necessary to use the largest of the values of thermal impedance $R_{th/ja}$ between chip and ambient. It occurs in the center of the chip and is calculated from the sum of $R_{th/cu-a}$ and the transfer impedance between silicon and copper $R_{th/j-cu}$ at this point.

At a distance of 0.5 cm from the chip center or greater, the dependence of the $R_{th/cu-a}$ curves on the chip area becomes negligible for the range of sizes examined (≤ 121 mm^2). In this case, dependence of the coupling impedance on chip area does not apply.

From **Figure 5** it can be seen that there is a rapid decrease in the magnitude of the coupling impedances in the immediate vicinity of the chip edges. This shows that the coupling effects need to be taken into account in the thermal design of the structure especially in cases of high packing density and high dissipation.

Figure 6 shows the dynamic response of the chip temperatures for a 50-A three-phase bridge if power is impressed in diode 2. The simulation assumed convectional heat transfer of 0.5 W/Kcm^2 at the base plate.

Diode 2 reaches the highest end temperature because of its internal heat generation. The highest degree of coupling occurs between chips mounted on the same copper surface. Because of the relatively poor thermal conductivity of the ceramic material compared to copper, the coupling impedances in the area between the copper surfaces are greatly reduced. The coupling temperature for IGBT 2 is therefore higher than that of the chips on the adjacent surface (diode 3 and IGBT 3).

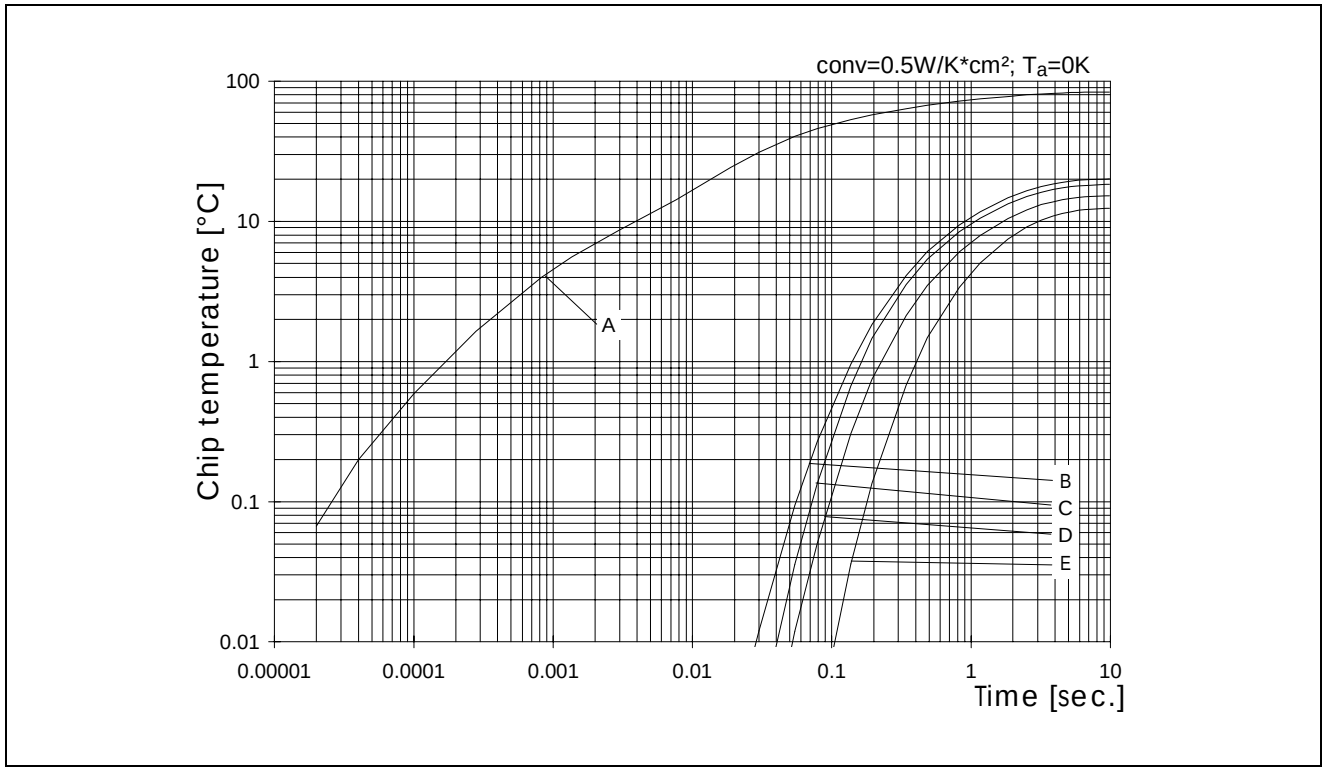

Figure 6
Chip Temperature Responses, with Power Impressed in Diode 2
(Module: BSM50GD120DN2; $\overline{P}_{L/Diode\ 2}$ = 118 W) (A-Diode 2 (x = 0, y = 0.9 cm); B-IGBT 2 (x = y = 0); C-Case (Diode2); D-Diode3 (x = 1 cm, y = 0.9 cm); E-IGBT 3 (x = 1 cm, y = 0))

An interesting aspect is the temperature variation with time at the base plate below diode 2 (curve 2). The lateral flow of heat on a copper surface is faster than the vertical flow towards the base plate. The thermal current reaches IGBT 2 before it reaches the base plate. The chips on the adjacent copper surface warm up later than the base plate. This indicates that the thermal coupling path to these chips runs via the lower layers and then vertically upwards, rather than laterally, near the surface, along the upper substrate layers.

The time constants for warming up the adjacent chips are by a factor of 10...100 greater than those for the self-heating of diode 2. The time-varying change of the coupling temperatures is, therefore, negligible compared to the self-heating effect and the temperature increase due to thermal coupling can be considered constant.

2 Thermal Design of PWM-Inverters

2.1 Dissipation Budget

Inverter configurations with DC link can be reduced to the half-bridge structure shown in **Figure 7**.

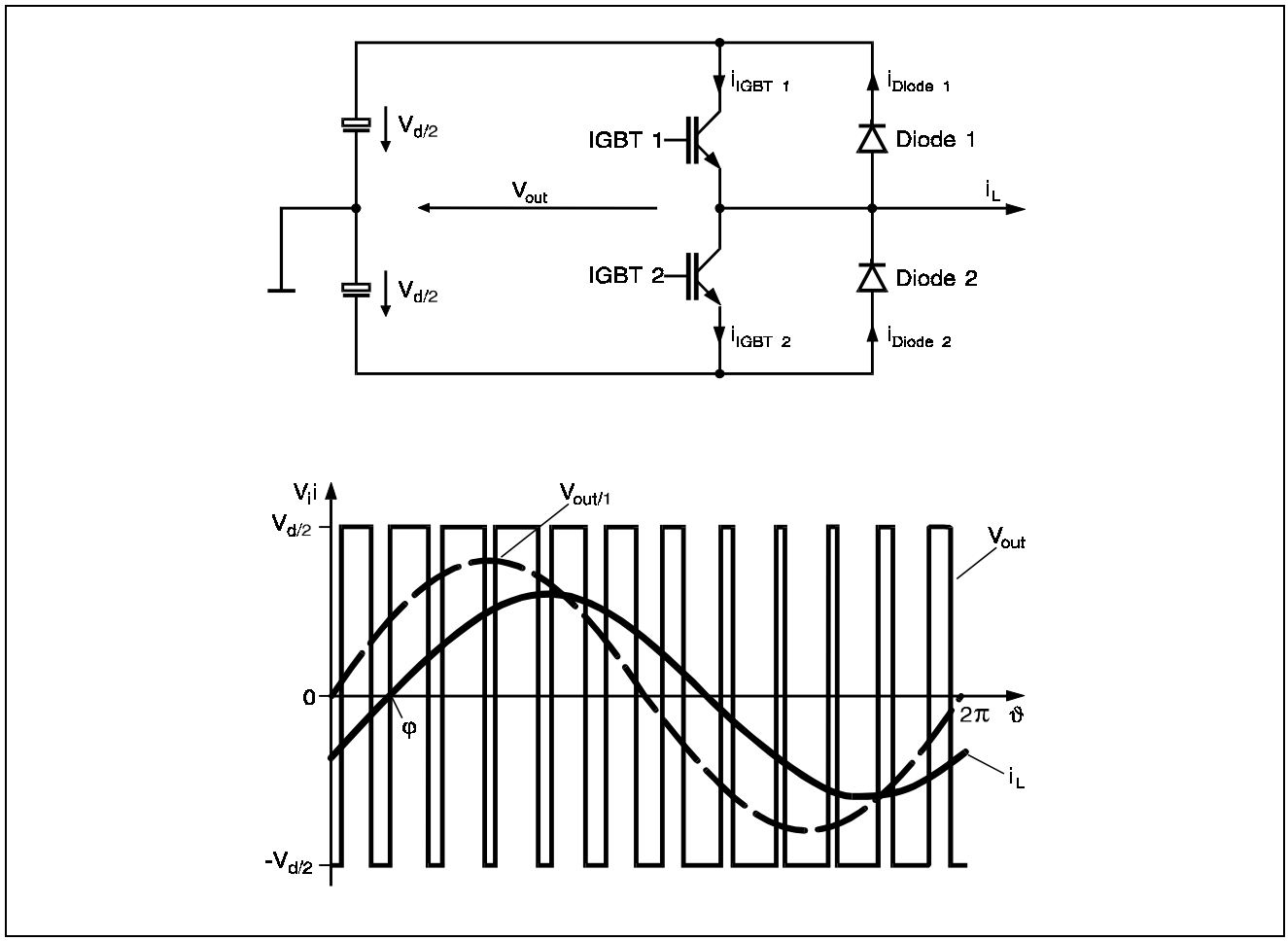

Figure 7
Basic Configuration for Calculating Dissipation Losses

IGBT 1 and IGBT 2 switch alternately. The control signals are generated via pulse width modulation. The resulting output signal consists of rectangular voltage pulses of different width (V_{out}). In the case of resistive-inductive loads, a sinusoidal output current i_L of frequency f_{out} flows which is shifted in phase relative to the fundamental of the output voltage ($V_{out/1}$).

A frequently used method for determining dissipation losses in configurations with sinusoidal output current has been described by Srajber. The dissipation losses are determined by calculation from a system of mathematical equations. To obtain this system of equations, the switching loss energies and the on-state voltages for IGBT and diode are measured as a function of the switch current in a DC chopper converter configuration at a given chip temperature and link circuit voltage, and are then approximated by polynomial equations.

By introducing temperature and voltage coefficients, these equations can be scaled to different chip temperatures and link circuit voltages.

This procedure yields equations 5 - 8 for the mean dissipation losses of IGBT and diode.

$\bar{P}_{Lfw/Diode} = A\ (I_{Lrms}) - m \times \cos\phi \times B\ (I_{Lrms})$ **equation 5**

$\bar{P}_{Lfw/IGBT} = C\ (I_{Lrms}) + m \times \cos\phi \times D\ (I_{Lrms})$ **equation 6**

$\bar{P}_{Lsw/Diode} = f_{sw} \times E\ (I_{Lrms})$ **equation 7**

$\bar{P}_{Lsw/IGBT} = f_{sw} \times F\ (I_{Lrms})$ **equation 8**

Equations 5 and 6 describe the on-state losses for IGBT and diode. The modulation factor m = 0...1 is the ratio of the amplitude of the output voltage to half the link circuit voltage. It is a measure for the voltage efficiency of the inverter. The phase shift between the output current and the fundamental of the output voltage is taken account of in the on-state losses by the power factor $\cos\varphi$ = – 1...1.

In the case of the on-state losses, the product $m \times \cos\varphi$ varies depending on the operating point of the inverter and causes the losses in diode and IGBT to vary. Consequently, this product is of prime importance for estimating the on-state dissipation in diode and IGBT at a given output current.

The operating point for maximum on-state dissipation is given for the IGBT by $m \times \cos\varphi \rightarrow -1$. The duty factor of the IGBT is at maximum here because of the high voltage efficiency. An increase in the power factor means that the maximum on-times of the IGBT and the current peak move closer together in time. In the limit case ($\cos\varphi = 1$), the IGBT is switched on for the longest time at maximum current. The diode exhibits maximum dissipation at the operating point where $\cos\varphi \rightarrow -1$. The negative power factor indicates that the inverter operates as a source. The flow of power is reversed. As $\cos\varphi$ decreases, the current flow through the diode increases.

The mean switching losses for the diode (equation 7) and IGBT (equation 8) are directly proportional to the switching frequency f_{sw}.

2.2 Determining the Chip Temperatures

2.2.1 Model Describing the Thermal System

The static and dynamic behavior of a power module is given by equation 2. On the basis of this equation, the equivalent circuit shown in **Figure 8** can be drawn up.

Components 1 - 4 represent the transfer impedance junction to case, including the heat-conducting paste (30 µm). C_5 and R_5 represent the heat sink. In this series model, the thermal resistances and capacities are determined by the time constants of the exponential functions (equation 2). They are imaginary and do not correspond to the values of the individual layers in the module structure. Scaling for different chip areas, with a given substrate configuration, can be performed using equation 3. The instantaneous dissipation losses calculated from the system of equations are impressed into the circuit via the current source.

Voltage sources are used for setting the ambient temperature T_{amb} and for taking account of the thermal coupling $\overline{T}_{coup}$. The calculation of $\overline{T}_{coup}$ is carried out in accordance with the method described in the previous chapter.

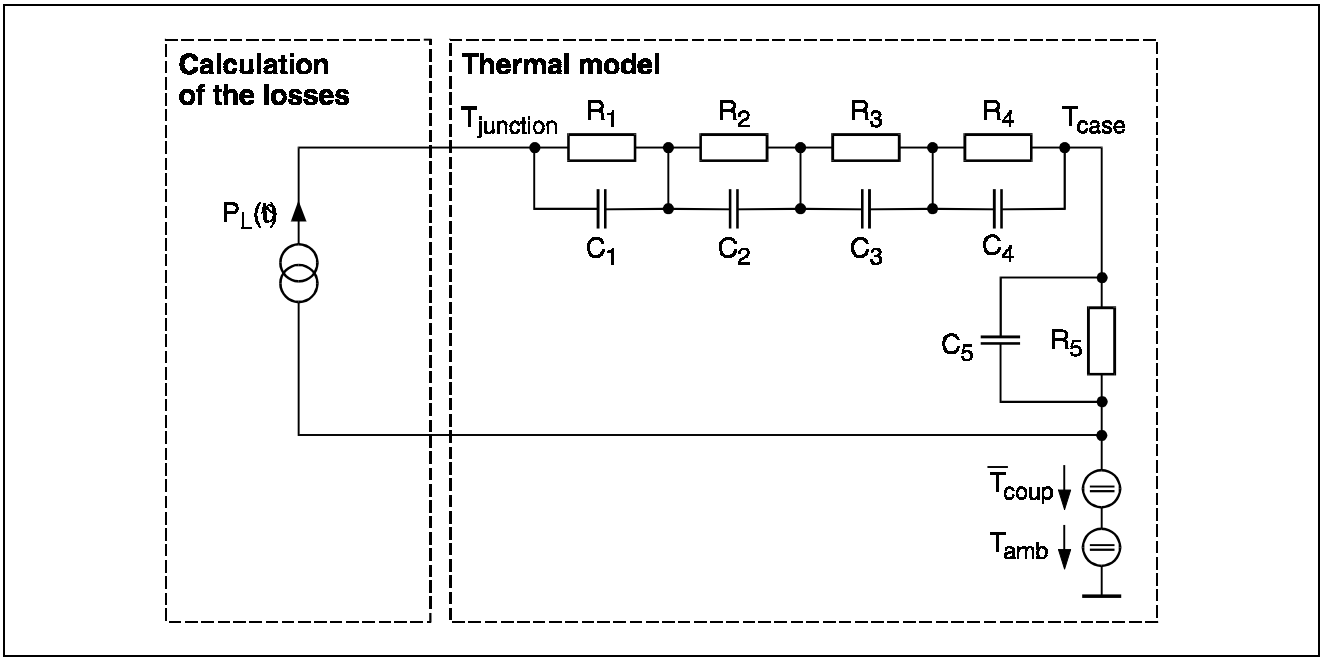

Figure 8
Equivalent Circuit for Simulating the Chip Temperatures

For dimensioning the cooling system, the hottest point in the module must be used as the starting point. The heat sinking system must be dimensioned for diode 2 and IGBT 2 (these have the highest temperature increase due to thermal coupling). The coupling temperature $\overline{T}_{coup}$ must accordingly be calculated for these two chips, taking into account all the adjacent chips.

2.2.2 Model for Simulating Chip Temperatures

A suitable method for determining the chip temperature responses is simulation using PSPICE.

The equivalent circuit described in the preceding section is used for this purpose. The simulation procedure is shown in **Figure 9**.

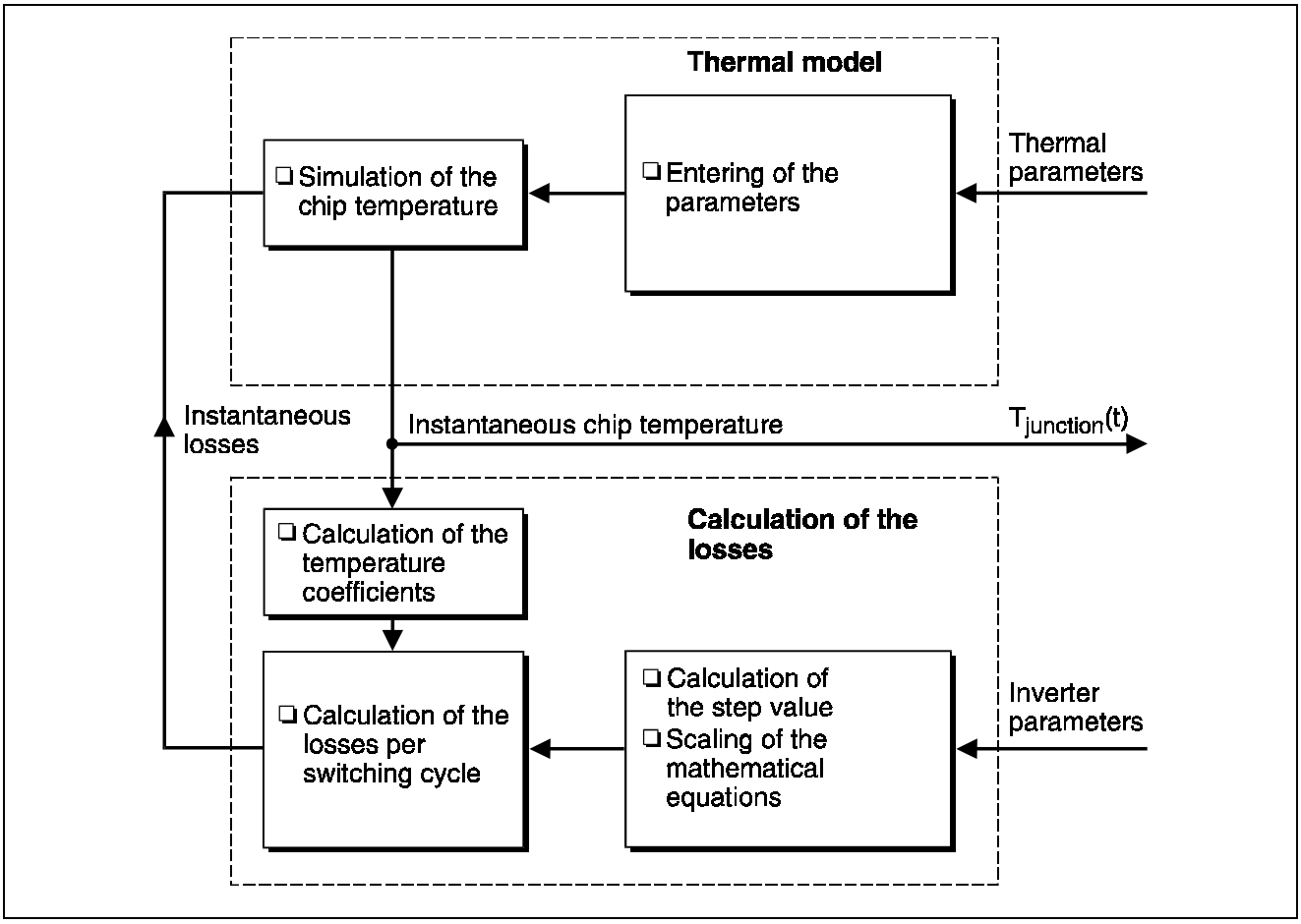

Figure 9
Model for Simulation of the Chip Temperatures

The calculation of chip temperature is performed iteratively for each pulse period, in accordance with the mathematical equations. At the beginning of each pulse period, the instantaneous chip temperature $T_j(t)$ is determined. With the aid of the temperature coefficients and allowing for $T_j(t)$, the dissipation for the following period is calculated. This dissipation is impressed into the RC network in the form of a current and again results in a change in the chip temperature, which then serves as the basis for the next calculation. This way of proceeding takes into account the influence of the chip temperatures on the dissipation budget.

The input parameters are the thermal parameters for defining the thermal equivalent circuit, and the inverter parameters for defining the converter operating point.

2.3 Design Criteria for the Use of Power Modules in PWM-Inverters

The transient thermal impedance changes with the response in time of the dissipation as shown in **Figure 2**. For a given inverter operating point, the dissipation response is determined by the output frequency f_{out}. The maximum value of impedance is obtained at an output frequency of $f_{out} \rightarrow 0$ Hz (static thermal resistance). The effective thermal impedance of the device diminishes as the output frequency increases. This leads to a reduction in maximum chip temperatures. **Figure 10** illustrates this behavior by showing the simulation results for the IGBT 2 at different output frequencies.

The response in time of the dissipation as a function of the current conduction angle $P_L(\vartheta)$, and consequently the mean dissipation $\overline{P}_L$, is the same for all simulated responses. The chip temperatures oscillate at the output frequency around the mean chip temperature T_j, which can be calculated from the mean dissipation and the static thermal resistance (equation 9). The temperature increase due to thermal coupling results in an increase of the mean chip temperature, in accordance with the equivalent circuit used. In the case of IGBT 2, this is only 3 K at the simulated operating point.

$$\overline{T}_j = R_{th/j-h} \times \overline{P}_L + T_h \qquad \textbf{equation 9}$$

Figure 10 shows particularly clearly the drastic reduction of the maximum chip temperatures occurring with only a slight increase of the output frequency, compared to the static case ($f_{out} \rightarrow 0$ Hz). Even at an output frequency of only 5 Hz, the maximum chip temperature is, in the simulated case, reduced by 20% compared to the worst case. At f_{out} = 50 Hz (nominal operating frequency), the maximum chip temperature is only 3 K above the mean temperature. This means that chip temperatures for $f_{out} \geq 50$ Hz can be determined via the average chip temperature.

The starting-up of a motor involves very low output frequencies. Depending on the size of the motor and the control method, these frequencies are in the order of the slip frequency (3-5 Hz). The high startup torque is achieved via a high startup current. This means that the dissipation losses in IGBT and diode are relatively high compared to normal operation. Consequently, because of the simultaneous existence of low frequencies and high dissipation, the startup phase is in most applications the worst-case condition and determines the thermal design of the cooling system.

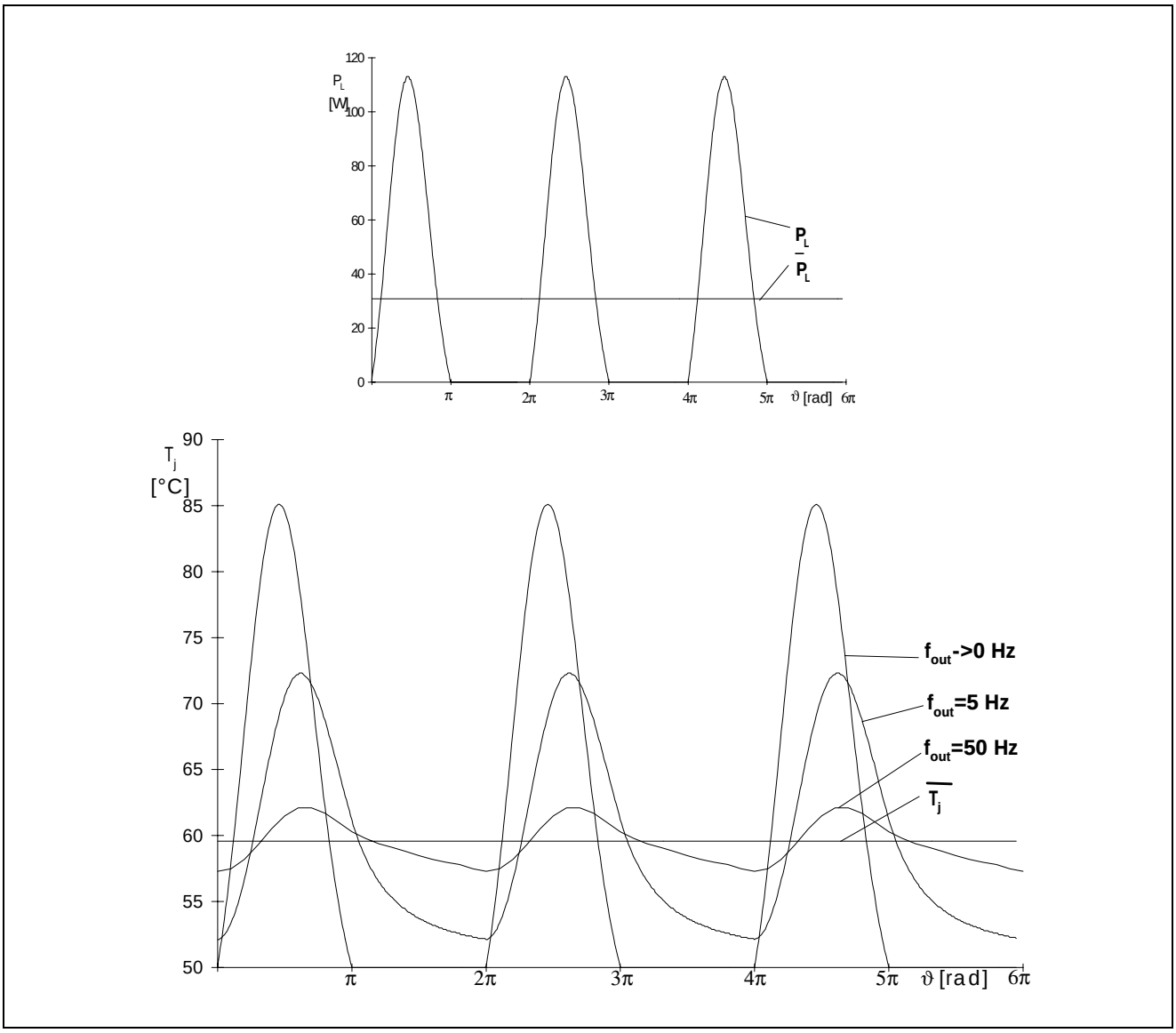

Figure 10
Variation of Maximum Chip Temperature with Output Frequency
(Module: BSM50GD120DN2; V_{CC} = 540 V; I_{Lrms} = 25 A; f_{sw} = 4 kHz; cosφ = 0.8; m = 0.8)

In inverter operation, all chips contribute to the dissipation budget. Because of the lateral spreading of the heat flow and the mounting of all chips in the same package, a homogenous, static temperature distribution can be assumed to exist at the contact surface between heat sink and module base in the case of the Sixpack module.

The thermal time constants of the heatsinks are several orders of magnitude higher than those of the modules (by a factor of 100 - 1000). Dynamic variation of the heatsink temperature can therefore be neglected down to very low frequencies. To examine the influence of the inverter output frequency, the variation of heatsink temperature T_h was measured underneath the chips during operation of the inverter. The maximum temperature difference between the individual measured points was 1 K; at frequencies down to f_{out} = 0.1 Hz only small amplitudes of variation were observed (≤ 0.5 K). The thermal impedance of the heatsink can therefore be calculated in accordance with equation 10.

$$R_{\text{th/h-a}} = \frac{T_{\text{h}} - T_{\text{a}}}{\sum \overline{P}_{\text{L}}}$$ **equation 10**

For the design of an inverter, the user has single switch, half-bridge and three-phase bridge designs at his disposal, depending on the required power range. To minimize system costs and size, the user endevours to use the power module with the highest level of integration. In the power range 50 A/1200 V, Siemens offers a single-package three-phase bridge (BSM50 GD120DN2) in ECONOPACK 2, in addition to the conventional half-bridge solution (BSM50 GB120DN2). Using the Sixpack, the designer achieves a saving of contact area of 45%. Another point in favour of using this module is the flow-solderability of the terminals. The time-consuming operation of making screw connections to the power terminals does not apply. This reduces the time for mounting a device by about 90%.

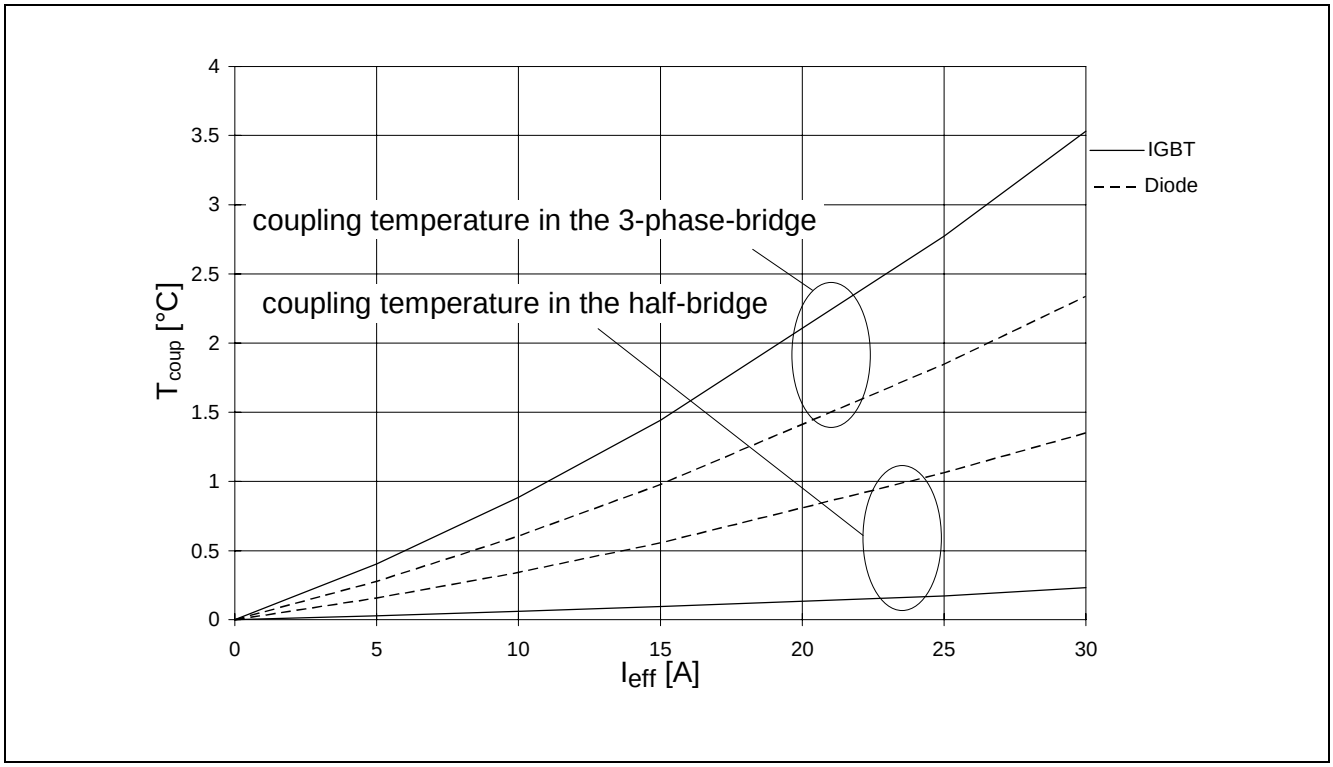

Figure 11
Comparison of Coupling Temperatures in Half-bridge and Sixpack Modules 1200 V/50 A (V_{CC} = 540 V; f_{sw}= 4 kHz; cosφ = 0.8; m = 0.8)

As pointed out before, the temperature rise due to mutual thermal coupling of the chips increases with increasing power density in the module. **Figure 11** compares the coupling temperatures for the half-bridge and Sixpack modules. The situation of the chips for the Sixpack can be seen from **Figure 6**. The expected temperature increase does occur, but even at very high output powers, the increase of around 3.5 K does not put any restrictions on the utilization of the semiconductors. The differences in coupling temperature of IGBT and diode result from differences in the positioning of the chips in the modules.