

High Power High Efficiency Multiple-Beam Klystron Design

Christopher James Lingwood BSc

A Thesis submitted for the Degree of Doctor of Philosophy

April 2010

For Hannah

(not that she wants it)

Christopher James Lingwood BSc

High Power High Efficiency Multiple-beam Klystron Design

A Thesis submitted for the Degree of Doctor of Philosophy April 2010

Abstract

This thesis investigated the design decisions and associated optimisation methods of a 1.3 GHz Multiple Beam Klystron (MBK) for use in the Compact Linear Collider (CLIC). In this regard refinements have been made both to the MBK design, and investigation and optimisation methods used. The high desired efficiency of 80% requires low perveance beams, to achieve the specification output power 20 beams are needed.

The choice of cavity used in the interaction structure of a klystron has a large impact on its size and efficiency. To optimise this a number of possible cavity designs were produced and compared to confirm selection of the most appropriate. The fundamental mode ($TM_{0,1,0}$) coaxial cavity was selected due to its superior R/Q of 130-210 Ω and suitability as a 2nd harmonic cavity. Although the dipole mode proved to be close in frequency to the operating mode (within ≈ 50 MHz), raising concerns of stability issues in an MBK. A novel model was developed using standard wake field theory to investigate the effects of this mode the klystron's stability. A strategy for shifting this mode using a coupled shifting gap was proposed and achieves a shift of ≈ 125 MHz, although the models findings suggest it is not a significant problem.

Existing methods of calculating dipole and higher order modes proved time consuming thus impeded a fully investigation of stability issues. An extended method of moments model allows efficient calculation of monopole and higher order modes. The model's basis functions are altered to represent a range of TM and TE modes with azimuthal variation, allowing their rapid and accurate calculation.

Optimising the klystron interaction structure by hand to find a viable configuration revealed shortcomings in this standard approach, although the target

efficiency was achieved. An algorithmic approach was deemed necessary to allow a full investigation within reasonable time limits. The field of evolutionary algorithms is presented and an evolutionary algorithm to automate the optimisation of klystron interaction structures was developed. A number of important related issues were dealt with and suitable interaction structures (optimised for efficiency, bandwidth, length and electron exit velocity) produced.

Finally a design was proposed for both the input and output couplers which is inspired by a coupler used in a gyrotron. Unconventionally, the latter exits the tube axially avoiding the focusing solenoid, but excessive heating may preclude its use.

Acknowledgements

I would first like to thank my wife Hannah for putting up with me throughout this work and making me many cups of tea.

Special thanks are due for the support Prof. Richard Carter and Dr. Graeme Burt for their patient assistance and input, without whom this thesis would not have been completed.

I would like to express my gratitude for the financial support and guidance of Thales Electron Devices, particularly Phillipe Thouvenin and Rodolphe Marchesin. Also to Erk Jensen and Igor Syrathev of Cern for their advice and enthusiasm.

I am grateful to Aaron Jensen for the kind provision of AJDisk, Kester Gunn for invaluable contribution in the development of the evolutionary algorithm and Jonathan Smith for his assistance with the Condor Pool.

Thanks are also due to my parents for their encouragement and grammatical assistance.

Nomenclature

Klystron Nomenclature

Tube properties

f Operating Frequency also represented as ω

N Number of beams

P_{in} RF input power

P_{out} RF output power

P_{av} Average RF output power

t_{pulse} Pulse repetition frequency

f_{pulse} RF pulse width

D Duty cycle (%)

η_t Total efficiency

η_e Electronic efficiency

G Gain at saturation

Beam properties

V_b Beam voltage

P_b DC beam power

I_{b0} DC Beam current

I_{b1} First harmonic of the beam current

a Tunnel radius

b Beam radius

$\frac{b}{a}$ Beam filling factor

u_0 Electron velocity

γ Lorentz factor of the electrons

Space charge wave properties

β_e Propagation constant

λ_0 Free space wavelength

ω_p Plasma frequency

λ_q Reduced plasma wavelength

R Plasma frequency reduction factor

ω_q Reduced plasma frequency

Gap properties

g Half gap length

V_g Gap voltage

E_g Gap electric field

I_g Current driven in equivalent circuit for the cavity by gap voltage

R_g Gap shunt impedance

R/Q R over Q

M_θ Gap coupling factor (transit angle)

M_r Radial gap coupling factor

M Total gap coupling factor

Cavity properties

Q_0 Q factor

Q_e External Q factor

W Stored energy

Gun properties

K Perveance

j_c Cathode current density

r_c Cathode radius

conv Area Convergence

Focus properties

B_b Brillouin focusing

B_b Brillouin ratio

B_0 Focusing Field

Evolutionary Algorithm Nomenclature

G Number of generations

Populations

P Population

$|\mathbf{P}|$ Population size

Decision space

N Number of decision values

X Individual

\mathbf{x} Decision vector of member of **P**

Solution space

M Number of solution values

\mathbf{z} Solution vector of member of \mathbf{P}

f fitness

γ score

Superscript

(g) at generation

* Optimum

Subscripts

(i) i^{th} individual

(m) Index of mother (m^{th} individual)

p relating to parents

s relating to Spartans

ϕ relating to offspring

\mathbf{F} Evaluation function

Strengths

σ_x search strength

σ_s kill strength

σ_μ mutation strength

$P(\mu)$ Probability of mutation

r Gaussian random number (0 mean and standard deviation of 1)

Significance

ι Significance index

Operators

$[\omega]$ Selection

$[f]$ Fitness

$[s]$ Spartan

$[\mu]$ Mutation

$[R]$ Recombination

Reproduction

p number of parents

s number of Spartans

\bar{x}_p mean of the parents

Notation structure

$\mathbf{P}_{subscript}^{(g)}$ Population notation structure

$property_{subscript(index)}$ general structure where *property* = for example *x*
or *f*

Method of Moments Nomenclature

h_n H field fourier coefficient

c speed of light

e_n E field fourier coefficient

μ permeability of free space

$\gamma = k_c = \sqrt{k^2 - k_z^2}$

ω angular frequency

k_z wave number in the guide

z longitudinal distance

β discrete values of k_z

r radial distance

E_z electric field maximum in the z
direction

n longitudinal variation index

k free space wave number ω/c

a_n amplitude of term in series

A constant

$G(k)$ Relationship between h_n and
 e_n

B constant

H_z magnetic field maximum in the
 z direction

i azimuthal variation index

ϵ permittivity of free space

E_θ electric field in the θ direction

Contents

Abstract	ii
Acknowledgements	iv
Nomenclature	v
Klystron Nomenclature	v
Evolutionary Algorithm Nomenclature	vi
Method of Moments Nomenclature	viii
Contents	ix
List of Figures	xix
List of Tables	xxxii
1 Literature Review	1
1.1 Introduction	1
1.2 Klystron Operation	5
1.3 Components of a Klystron	7
1.3.1 Electron Gun	7
1.3.1.1 Thermionic Emission	9
1.3.1.2 Cathode Chemistry	10
1.3.1.3 Cathode Lifetime	11
1.3.1.4 Cathode Operating Regimes	12
1.3.1.5 Perveance	13

1.3.2	Beam Confinement	14
1.3.2.1	Solenoid Focusing	16
1.3.2.2	PPM Focusing	17
1.3.3	Output	18
1.3.4	Windows	19
1.3.5	Limiting Factors	21
1.3.5.1	Voltage Breakdown	21
1.3.5.2	Multipactor	21
1.3.5.3	Monotron Oscillations	22
1.3.6	Examples	23
1.3.7	Collector	26
1.4	Multiple Beam Klystron (MBK)	28
1.4.1	MBK Focusing	29
1.4.2	Choice of Number of Beamlets	32
1.4.3	MBK Configurations	34
1.4.3.1	$TM_{0,1,0}$ Fundamental Cavity	35
1.4.3.2	$TM_{0,2,0}$ Cavity	36
1.4.3.3	$TM_{0,1,0}$ Coaxial Cavity	37
1.4.3.4	Higher Order Mode MBK	37
1.4.3.5	$TM_{m,1,0}$ Coaxial Cavity	38
1.4.3.6	$TM_{m,1,0}$ Cylindrical Cavity	39
1.5	Klystron Analysis	40
1.5.1	Basic Tube Properties	40
1.5.2	Small Signal	42
1.5.3	Large Signal	44
1.6	Klystron Specification	45
1.7	Methodology	46
2	Cavity Design	48
2.1	Introduction	48

2.2	Methodology	50
2.2.1	Modelling Fundamental Mode	50
2.2.2	HOM Calculation - Sector Method	52
2.2.3	MBK R/Q Definition	53
2.3	Re-entrant Cavity ($TM_{0,1,0}$)	56
2.3.1	Description	56
2.3.2	Initial Parameter Results	57
2.4	Recessed Re-entrant Cavity ($TM_{0,1,0}$)	58
2.4.1	Description	58
2.4.2	Initial Parameter Results	59
2.5	Fundamental Mode Coaxial Cavity ($TM_{0,1,0}$)	62
2.5.1	Description	62
2.5.2	Initial Parameter Results	63
2.6	HOM Coaxial Cavity ($TM_{10,1,0}$)	65
2.6.1	Description	65
2.6.2	Initial Parameter Results	65
2.7	Whispering Gallery Cavity ($TM_{10,1,0}$)	67
2.7.1	Description	67
2.7.2	Initial Parameter Results	68
2.8	Preliminary Conclusions	69
2.9	Revised Parameters	70
2.9.1	Cavity Gap Length	71
2.9.2	Radius of Cathode Circle (r_{pcr})	72
2.9.3	Area Convergence	73
2.9.4	Tunnel to Beam Ratio (b/a)	75
2.9.5	Beam Ridge Width (R_0)	75
2.9.6	Final Parameters	75
2.10	Recessed Re-entrant Cavity ($TM_{0,1,0}$)	77
2.10.1	Revised Parameter Results	77

2.11	Fundamental Mode Coaxial Cavity ($TM_{0,1,0}$)	81
2.11.1	Revised Dimensions	81
2.11.2	HOM in Coaxial Cavity	84
2.11.3	2 nd Harmonic Cavity	86
2.12	Conclusions	87
3	Dipole Mode	89
3.1	Introduction	89
3.2	Effect of r_0 on Dipole Mode	90
3.3	Waveguide Theory	91
3.3.1	Theory Verification	93
3.3.2	Results	96
3.4	Shifting Gap	98
3.4.1	Dipole Mode Suppression	98
3.4.2	Theory	99
3.4.3	Results	100
3.4.4	Central Post	103
3.5	Conclusions	105
4	Method Of Moments	106
4.1	Introduction	106
4.1.1	Method of Moments	107
4.2	TM Mode Derivation	109
4.2.1	Field Component Equations	109
4.2.2	Application of Boundary Conditions	111
4.2.3	Convert Between E and H (\mathbf{G}^I and \mathbf{G}^{II})	113
4.2.4	Continuity Conditions (\mathbf{P} and \mathbf{Q})	115
4.2.4.1	E-field Continuity \mathbf{P}	116
4.2.4.2	H-field Continuity \mathbf{Q}	117
4.2.5	Finding the Resonant Modes	118

4.2.6	Implementation	119
4.3	<i>TM</i> Mode Results	121
4.3.1	Azimuthal Variation Index (i) 0	122
4.3.2	Azimuthal Variation Index (i) 1	123
4.3.3	Inaccurate modes	124
4.4	<i>TE</i> Modes Derivation	124
4.4.1	Field Component Equations	125
4.4.2	Apply Boundary Conditions	126
4.4.3	Convert between E and H (\mathbf{G}^I and \mathbf{G}^{II})	127
4.4.4	Continuity Conditions (\mathbf{P} & \mathbf{Q})	127
4.4.4.1	H-field Continuity \mathbf{Q}	128
4.4.4.2	E-field Continuity \mathbf{P}	129
4.4.5	Finding the Resonant Modes	130
4.5	<i>TE</i> Mode Results	131
4.5.1	Azimuthal Variation Index (i) 1	131
4.6	Coupled Modes Derivation	132
4.6.1	Mixing E_θ	132
4.6.2	Mixing H_θ	134
4.6.3	Convert Between θ and z (H & J)	135
4.6.4	Recap	135
4.6.5	Solution	136
4.7	Mixed Mode Results	138
4.7.1	Azimuthal Variation Index (i) 0	138
4.7.2	Azimuthal Variation Index (i) 1	139
4.8	Conclusions	140
5	Wake-field Stability Analysis	142
5.1	Introduction	142
5.2	Time domain	144
5.2.1	Theory	144

5.2.1.1	Beam Current (I_b) Variation	145
5.2.1.2	Wake-fields	145
5.2.2	Results	148
5.2.2.1	Frequency of Dipole Mode	148
5.2.2.2	Dipole Q factor	149
5.2.2.3	Effect of Beam Current (I_b) Variation	151
5.2.2.4	Loss of Beams	152
5.3	Conclusions	154
6	Interaction Structure	156
6.1	Introduction	156
6.2	Manual Optimisation	158
6.2.1	High R/Q Interaction Structures	160
6.2.1.1	Initial Parameters	160
6.2.1.2	Revised Parameters	162
6.2.1.3	Magnetic Field	165
6.2.2	Low R/Q Interaction Structures	167
6.2.2.1	Initial Parameters	167
6.3	Conclusion	168
7	Literature Review Evolutionary Algorithm	170
7.1	Introduction	170
7.2	What is an EA (Evolutionary Algorithm)?	172
7.3	Selection	174
7.3.1	Tournament Selection	176
7.3.2	Truncation Selection	177
7.3.3	Linear Ranking	178
7.3.4	Exponential Ranking	180
7.3.5	Loss of Diversity	181
7.4	Recombination	182

7.4.1	PCX (Parent Centric Recombination)	183
7.4.2	MrCX (Mother Centric Recombination)	186
7.5	Population Management	189
7.6	Multi-Objective EAs	190
7.6.1	Classical Methods	191
7.6.1.1	Method of Objective Weighting	191
7.6.1.2	Method of Distance Functions	192
7.6.1.3	Min-Max Formulation	192
7.6.1.4	Drawbacks of Classical Methods	193
7.6.2	Domination	194
7.6.3	Clustering (k^{th} Nearest)	195
7.6.4	VEGA (Vector Evaluated Genetic Algorithm)	195
7.6.5	NSGA (Non-dominated Sorting Genetic Algorithm)	196
7.6.6	NSGAI (Non-dominated Sorting Genetic Algorithm II)	199
7.6.7	SPEA (Strength Pareto Evolutionary Algorithm)	201
7.6.8	SPEA2 (Strength Pareto Evolutionary Algorithm 2)	203
8	Evolutionary Algorithm	207
8.1	Klystron Optimisation	207
8.2	Klystron Simulation Issues	208
8.2.1	Pareto Search Pressure Focusing (PSPF)	208
8.2.2	Simulation Vetoes	210
8.2.3	Bandwidth Targeting	211
8.2.4	Reflected Electrons	212
8.3	Algorithm Refinements and Innovations	213
8.3.1	Spartan Operator	214
8.3.2	Population Management	216
8.3.2.1	Rejection	216
8.3.2.2	Significance	217
8.3.2.3	Significance Population Management	218

8.4	Simulation Tools	218
8.4.1	High Throughput Computing Resources (CONDOR) . . .	218
8.4.1.1	Asynchronous Calculation	219
8.4.2	AJDisk	220
8.4.2.1	Disk/Ring Model	221
8.4.2.2	Equations of Motion	221
8.4.2.3	Cavity Fields	222
8.4.2.4	Space Charge	223
8.4.2.5	Calculation Structure	225
8.4.2.6	Convergence	225
8.5	Methodology and Calibration	228
8.5.1	SpartanEA Implementation	228
8.5.2	Test Problems	229
8.5.2.1	Ellipsoidal Problem	229
8.5.2.2	Zitzler T_6	230
8.5.3	Performance Metrics	231
8.5.3.1	Domination Metric	231
8.5.4	EA Setup	232
8.5.5	Result Presentation	233
8.5.6	Search Strength	234
8.5.7	Mutation	236
8.5.8	Number of Parents	237
8.5.9	Population Calibration	239
8.5.10	Spartans	240
8.5.11	Population Management	241
8.5.12	Calibration Conclusions	242
8.6	Klystron Optimisation Results	243
8.6.1	TH2178 Comparison	244
8.6.1.1	TH2178 Specification	244

8.6.1.2	TH2178 AJDisk Calibration	245
8.6.1.3	TH2178 Optimisation	247
8.6.2	Improved TESLA Optimisation	254
8.6.2.1	Improved TESLA Specification	254
8.6.2.2	Improved TESLA AJDisk Calibration	256
8.6.2.3	Improved TESLA AJDisk Optimisation	258
8.7	Conclusions	265
9	External Coupling	267
9.1	Introduction	267
9.2	Coaxial Output Coupler	268
9.2.1	Convergence	271
9.2.2	Simulations	272
9.3	Input coupler	275
9.3.1	Convergence	277
9.3.2	Simulations	278
9.4	Conclusions and Further Work	280
10	Conclusions and Further Work	281
10.1	Conclusions	281
10.2	Further work	288
A	MWS Modifications	290
A.1	Introduction	290
A.2	Polynomial Analysis	290
A.2.1	Implementation - 2 Free Dimensions	292
A.2.2	Accuracy	294
A.2.3	R/Q Prediction	295
A.2.4	Conclusions	295

B Method Of Moments	296
B.1 \mathbf{G}_n^{xTE} Function	296
B.2 \mathbf{H}^x and \mathbf{J}^x	296
B.3 Azimuthal - Variation Index (i) 2	297
B.3.1 TM Basis Functions	297
B.3.2 Mixed TE and TM Basis Functions	298
Bibliography	299

List of Figures

1.1	Tunnel Configuration of a direct acceleration (TESLA) and TBA (CLIC) accelerator showing the relative complexities	3
1.2	Overall layout of the CLIC complex at 3 TeV[1]	4
1.3	Single beam klystron schematic[2]	5
1.4	Applegate diagram showing convergence of electrons in time for a single RF phase[2]	6
1.5	Diagram showing the components of an electron gun. The control electrode is not used on high current cathodes.[2]	8
1.6	Cathode lifetime with respect to cathode loading[3]	12
1.7	Voltage vs. current for an electron gun showing the space charge and temperature limited regions	12
1.8	PPM focused beam [4]	18
1.9	View of output cavity with mini-windows[3]	20
1.10	Cross sectional view of E3746 klystron[5]	24
1.11	Cut through of a KEK klystron output cavity showing the double gap configuration [6]	25
1.12	Multiple gap TW output structure in XB72K[7]	25
1.13	Tilted output couplers for damping of TE ₀₁ modes[8]	26
1.14	Schematic of a multiple beam klystron[9]	28
1.15	Magnetic field shaping iron [10]	30
1.16	PPM focused gun[6]	31

1.17	Different ways of representing the relationship between micro-perveance and efficiency	33
1.18	Relationship between number of beams, beam current and gun voltage at constant power and perveance. The chosen combination of number of beams, voltage and current is marked.	34
1.19	E_z intensity of the $TM_{0,1,0}$ mode in the transverse plane	35
1.20	E_z intensity of the $TM_{0,2,0}$ mode in the transverse plane	36
1.21	E_z intensity of the $TM_{0,2,0}$ mode in the transverse plane of coaxial cavity	37
1.22	E_z intensity of the $TM_{m,1,0}$ mode in the transverse plane of coaxial cavity[3]	38
1.23	E_z intensity of the $TM_{m,1,0}$ mode in the transverse plane of coaxial cavity[3]	39
2.1	The use of offset grids to model the magnetic and electric fields[11]	50
2.2	The boundary conditions used for simulations performed on the re-entrant, recessed re-entrant, coaxial fundamental cavity cavities. Green: electric symmetry plane ($E_{\perp} = 0$) at (X,Y), Blue: magnetic symmetry plane ($H_{\perp} = 0$) at (Y,Z) & (X,Z)	51
2.3	The boundary conditions used in MWS for the coaxial HOM cavity, whispering gallery cavities (both $TM_{10,1,0}$ modes). Green: electric boundary, Blue: magnetic boundary All boundaries in this case are electric boundaries. The wedge shows the minimum amount of cavity that can be meshed while still representing the correct fields.	53
2.4	Simple model of interaction gap and beam with AC gap current I_g and AC voltage V_g ; cavity losses R_g and first harmonic of the beam current I_{b1}	54

2.5	4 beam klystron showing the stored energy interacted with by one beam (1/4 of the total energy). So the R/Q calculated for a single beam in a full cavity is in fact 1/N times the correct R/Q seen by the beams (due to the increased stored energy being associated with the beam).	55
2.6	Re-entrant cavity	56
2.7	Re-entrant cavity: E_z plot in radial plane of $TM_{0,1,0}$ Mode	57
2.8	Re-entrant Cavity: R/Q contours plotted against h & r_1 (plus 1.3 GHz frequency contour) showing maximum R/Q of $\sim 70 \Omega$	57
2.9	Recessed re-entrant cavity	59
2.10	Recessed re-entrant cavity: E_z plot in radial plane of $TM_{0,1,0}$ Mode	59
2.11	Recessed re-entrant cavity: Iso-frequency surface where $f=1.3$ GHz. R/Q is represented by the colour map, showing a maximum of $\sim 220 \Omega$	60
2.12	Recessed re-entrant cavity: R/Q contours plotted against h & r_1 (plus 1.3 GHz frequency contour) where $h_r = 101.77$ mm	61
2.13	Coaxial cavity	62
2.14	Fundamental mode coaxial: E_z plot in radial plane of $TM_{0,1,0}$ Mode	63
2.15	Fundamental mode coaxial: Iso-frequency surface at 1.3 GHz. R/Q is represented by the colour map	63
2.16	Fundamental mode coaxial: R/Q contours plotted against h & r_1 (plus 1.3 GHz frequency contour) for coaxial cavity ($r_{pcr} = 62.8$)	64
2.17	HOM coaxial cavity: E_z plot in radial plane of $TM_{10,1,0}$ mode	65
2.18	HOM coaxial cavity: Iso-frequency surface at 1.3 GHz. R/Q is represented by the colour map	66
2.19	HOM coaxial cavity: $r_{pcr} = 400$ mm	66
2.20	Whispering gallery cavity	67
2.21	Whispering gallery cavity: E_z plot in radial plane of $TM_{10,1,0}$ Mode shape in the whispering mode cavity	68

2.22	Whispering gallery cavity: Iso-frequency surface at 1.3 GHz. R/Q is represented by the colour map for whispering gallery cavity with beam ridge simulated in a segment. The gaps in the plot are an artifact of the interpolation methods used in MatLab.	69
2.23	$3.5 A/cm^2$ is a good compromise between lifetime and overall tube radius.	73
2.24	The definition of cathode loading allows the selection of the beam radius with consideration of the area convergence of the gun and the focusing field strength required.	74
2.25	A comparison between the new and old beam/cathode arrangements	76
2.26	Recessed re-entrant cavity (repeated from Sec. 2.4)	77
2.27	Frequency change of fundamental mode in GHz w.r.t. the height of the cavity. Increasing beyond its lower limit does not tend towards 1.3 GHz.	78
2.28	Frequency change of fundamental mode in GHz w.r.t. the outside radius of the cavity. Increasing beyond its lower limit does not tend towards 1.3 GHz.	79
2.29	Frequency change of fundamental mode in GHz w.r.t. the height of the recess. Increasing beyond its lower limit the frequency tends towards 1.3 GHz but not before the cavity becomes too large.	80
2.30	Coaxial Cavity (repeated from Sec. 2.5)	81
2.31	R/Q against h for coaxial cavities with frequency 1.3 GHz with varying R_0 . Increasing R_0 decreases maximum R/Q	83
2.32	Coaxial cavity maximum R/Q (Ω) vs R_0 (mm)	84
2.33	Coaxial cavity mode spectra up to 7 GHz showing R/Q and optimal R/Q and harmonics of 1.3 GHz. Of particular note is the mode very close to 1.3 GHz (a dipole mode).	85
2.34	Coaxial 2 nd harmonic cavity R/Q contours and 2.6 GHz contour .	87

3.1	Dipole mode of the fundamental mode coaxial cavity	89
3.2	A shift of only around 63 MHz can be expected and the R/Q of the dipole mode remains above the monopole mode. (see fig. 2.32)	90
3.3	The conversion between coaxial cavity and ridged waveguide . . .	91
3.4	k_z, ω plot showing discrete wavelength solutions where the periodic boundary conditions are satisfied showing the mode spacing in ω increases with increasing β	92
3.5	Plots of E_z for the two test series of modes in the coaxial cavity (in these cases $x = 0$)	94
3.6	Percentage deviation of the frequency prediction from the MWS results	94
3.7	Plots of E for the two positions of the TE_{x11} series of modes ($x = 0$ for these plots)	95
3.8	Percentage deviation of the frequency prediction from the MWS results	95
3.9	Percentage deviation of the frequency prediction from the MWS results	96
3.10	Ratio of monopole mode to dipole and quadrupole mode for various cavity sizes showing the CPI design[12] and the current design	97
3.11	Equivalent circuit representing the $TE_{0,1}$ in rectangular waveguide[13]	99
3.12	Proposed dipole shifting design shown in "solid" vacuum view .	100
3.13	Frequency against the size of the shifting gap in mm. As the gap increases from 0 the dipole mode moves away until it crosses the descending quadrupole mode. It is at this point that the optimum gap lies.	101
3.14	E_z plots of new modes which interact with the beam, now split by ≈ 125 MHz	102

3.15	The two new LOM (Lower Order Modes) which may be hard to damp. However the E-field is low in the beam region and their frequency is far from the operating frequency	103
3.16	Shifting gap design with a central post shown in “solid” vacuum view.	103
3.17	The transition between no central post (post radius = 0) and no shifting gap (post radius = r_1) (the termination of lines are due to the frequency range of the simulation). The original operating mode and dipole ($TM_{0,1,0}$, $TM_{1,1,0}$) can be seen to transition into the new LOMs and the new operating mode and dipole ($TM_{0,2,0}$, $TM_{1,2,0}$) transition into a HOM.	104
4.1	Representation of a re-entrant cavity used for the method of moments. Note the division into region <i>I</i> and region <i>II</i>	108
4.2	Three regions configuration	108
4.3	Eq. 4.45 displayed graphically, clearly showing solutions. These plots are for $i = 0$	121
4.4	Plots of E_z and H_θ calculated by the Method of Moments along the interface between the 2 regions	121
4.5	MAFIA plot of an inaccurate mode which is in fact a <i>TE</i> mode . . .	124
4.6	Recessed re-entrant cavity split into region <i>I</i> and region <i>II</i>	128
5.1	Normal probability distribution of percentage variation of I_b between the beams	145
5.2	Position of appropriate dipole modes (denoted by dot-dashed lines) and their relative strength of interaction with the beam . . .	149
5.3	The effect of the dipole Q_0 on the dipole mode content for the original and new dipole modes. Showing the increasing dipole content with <i>decreasing</i> dipole Q_0	150

5.4	The effect of the dipole Q_0 on the dipole mode content for the original and new dipole modes. Showing the decreasing dipole content for very small Q_0 s.	151
5.5	The effect of the variation in I_b between the beams. Showing that even a small variation creates an appreciable dipole component.	152
5.6	Effect of the loss of beams on the dipole content. The loss of just a single beam has a significant effect on dipole content. This effect decreases as number of beams lost increases.	153
5.7	Effect of the loss of beams on the dipole content. For a dipole frequency of 1.3 GHz the high R/Q leads to a positive dipole content where only one beam is operational.	153
6.1	Applegate diagram showing the perceived advantage of $\frac{1}{4}\lambda_q$ cavity spacing for at least the first cavity[2]	157
6.2	~80% efficient high R/Q (Recessed re-entrant and fundamental mode coaxial cavity) interaction structure at a power of 748 kW per beam (14.9 MW in total) with 14 W input power	161
6.3	High R/Q Applegate diagram relating to the above interaction structure (see fig. 6.2)	161
6.4	~80% efficient high R/Q (Recessed re-entrant and fundamental mode coaxial cavity) interaction structure using the revised parameters at a power of 755 kW per beam (15.1 MW in total) with 10 W input power	163
6.5	High R/Q Applegate diagram relating to the above interaction structure using revised parameters (see fig. 6.4)	164
6.6	Interaction structure transfer curve for a klystron using the high R/Q cavities with the revised parameters	165

6.7	~80% efficient high R/Q (Recessed re-entrant and fundamental mode coaxial cavity) interaction structure using the revised parameters at a power of 755 kW per beam (14.9 MW in total) with 10 W input power with efficiency and radial movement and magnetic field simulated	166
6.8	Particle tracks down the above klystron interaction structure with radial movement and magnetic field simulated showing the need for increased drift tube radius	166
6.9	~70% efficient low R/Q (HOM coaxial and whispering gallery cavity) interaction structure at a power of 677 kW per beam (13.5 MW in total) with 14 W input power	167
7.1	Basic Evolutionary Algorithm structure	174
7.2	Tournament selection	177
7.3	Truncation selection	178
7.4	Linear ranking	179
7.5	Exponential ranking	181
7.6	Dependence on the loss of diversity on selection intensity for Linear Ranking, Exponential Ranking and Tournament and Truncation selection. Bickel and Thiele [14]	182
7.7	Probability distributions of UNDX and SPX recombination operators	183
7.8	Properties of PCX[15]	184
7.9	Probability distribution of offspring (\mathbf{x}_ϕ) generated by PCX in one dimension (where $\bar{\mathbf{x}}_p$ is the mean of the parents and $\mathbf{x}_{p(m)}$ is the mother)	185
7.10	Probability contour at 1 standard deviation of offspring generated by PCX in two dimension with three parents ($\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$ and $\mathbf{x}^{(p)}$) $\bar{d}_p = \frac{d_{p(1)} + d_{p(2)}}{2}$ [15, 16]	186
7.11	Vector operations used in MrCX (to scale)[17](see alg. 7.6)	188

7.12	Representations of a multi-objective problem. Showing the advantage of a true multi-objective solution over a classical method.[18]	191
7.13	The effect of eq. (see eq. 7.14) on the space behind a series of points. The darker colours represent a more highly dominated area.	194
7.14	Solution of Shaffer's F1 (see eq. 7.17) after 500 generations using NSGA and VEGA, highlighting VEGA's tendency to fixate on particular solutions [19]	198
7.15	NSGAI procedure, where $P_t = \mathbf{P}_p$ and $Q_t = \mathbf{P}_\phi$ and $F_n = f_n$ [20]	199
7.16	Comparison between NSGAI and SPEA in a minimisation problem where SPEA finds the solution better but NSGAI defines the front better on Zitzler T6[21] (see sec. 8.5.2.2)[20]	201
7.17	NSGAI's fitness assignment (neglecting crowding). Colour groups are dominated by the same number of other solutions.	201
7.18	Fitness assignment in SPEA[22]	203
7.19	Comparison on the fitness assignment schemes of SPEA(left) and SPEA2(right). $R(i)$ is shown for SPEA2.[23]	204
7.20	Demonstration of the truncation used by SPEA2 where the target population size is 5. k^{th} nearest crowding (see sec. 7.6.3) is used to rank the individuals which helps preserve end points.[23] . . .	205
8.1	Relationship between length and efficiency	209
8.2	Demonstration of the effect of PSPF (Good→Bad) the arrows represent the search pressure (note the sign of the y axis)	210
8.3	A more suitable definition of bandwidth for use in an evolutionary algorithm	211
8.4	Optimisation of reflected electrons to avoid reflected electrons at the output gap (purple and green markings should be ignored). This is achieved by maintaining the exit velocity of the disks. . . .	213

8.5	Probability distribution for MrCX alone compared with the distribution for the Spartan operator[16]	214
8.6	Examples of spartan behaviour involving two Spartans acting on two offspring in turn (a and b)[16]	215
8.7	Schematic representation of asynchronous calculation	220
8.8	The two space charge representations used by AJDisk to model the beam	221
8.9	Flow chart showing AJDisk's operation, reproduced from [24]	225
8.10	Number of disks and rings against reported output power	226
8.11	Number of disks against <i>normalised</i> time. The time units used are comparable to seconds for a modest CPU.	227
8.12	A flow chart representing the flow of populations around the SpartanEA	229
8.13	Methods of evaluating the area dominated by a number of points	232
8.14	An example of a box plot (or box-and-whisker diagram) with an overlaid plot of the mean	234
8.15	Single objective search strength σ_x sweep performed on the ellipsoidal problem[16]	235
8.16	Multi-objective search strength σ_x sweep performed on T_6 . Showing that a search strength of 0.8 improves the dominated metric.	236
8.17	Multi-objective mutation strength σ_μ sweep performed on T_6 . Showing that mutation has at best an inconclusive effect on the dominated metric.	237
8.18	Multi-objective number of parents p sweep performed on T_6 . Showing that the original assumption of $N + 1$ holds.	238
8.19	Multi-objective initial population size $ \mathbf{P}^{(0)} $ sweep performed on T_6 . Showing that the optimal initial population is 300.	239
8.20	Single-objective number of spartan s sweep performed on T_6 . Showing that the optimal number of Spartans is around 40.	241

8.21	Comparison of population limitation models. This demonstrates there is a slight benefit in using rejection combined with “fitnificance”	242
8.22	The TH2178 klystron[25]	244
8.23	Modulation current plot growth for the fundamental and 2 nd harmonic frequencies along the length of the klystron	245
8.24	Calibration of AJDisk simulation using the gap coupling factor as the free parameter	246
8.25	Modulation current plot growth for the fundamental and 2 nd harmonic frequencies along the length of the klystron after calibration	246
8.26	Convergence of efficiency against iterations. The top line shows maximum value and the middle line the average.	248
8.27	Deviation of the drift lengths (in meters) from the specification drifts in the final population against efficiency. Thales design drifts are marked with a vertical dotted line. The colour map and point size represent u_{slow} , bigger is faster.	250
8.28	Deviation of the frequencies (in MHz) from the specification drifts in the final population against efficiency. Thales design frequencies are marked with a vertical dotted line. The colour map and point size represent u_{slow} , bigger is faster.	252
8.29	Plots of the relationships between the components of the solution vectors. Unless otherwise stated the colour map and point size represent u_{slow}	254
8.30	Modulation current plot growth along the length of the klystron .	256
8.31	Calibration of AJDisk for a range of input powers (specification $P_{in} = 10W$) over a range of beam sizes with optimal gap coupling factors	257
8.32	Comparison between AJDisk (black) and Klys2D (grey dashed) .	258

8.33	Convergence of efficiency against iterations. The top line shows maximum value and the middle line the average and the bottom the minimum.	259
8.34	The drift lengths (in meters) of the individuals in the final population vs efficiency. Hand optimised drifts are marked with a vertical dotted line. The colour map and point size represent u_{slow} , bigger is faster.	261
8.35	The frequencies (in MHz) of the individuals in the final population. Hand optimised frequencies are marked with a vertical dotted line. The colour map and point size represent u_{slow} , bigger is faster.	263
8.36	Plots of the relationships between the components of the solution vectors. Unless otherwise stated the colour map and point size represent u_{slow}	265
9.1	Gyro-klystron output coupler design used by Calabasas Creek Research, Inc.[26]	268
9.2	Electric fields of the modes supported by coaxial transmission line[27]	269
9.3	“Solid” vacuum representation of the output coupler with the ports marked	271
9.4	Convergence behaviour of the cavity	271
9.5	Diagrams of the fields in the output cavity and coupler for the operating mode	272
9.6	Plot of E_z around a circle passing through the beam centres. Showing the ripple caused by the conductive posts.	274
9.7	E_z of the dipole mode in the output cavity. It’s frequency remains within ≈ 50 MHz of the operating mode	275
9.8	“Solid” vacuum representation of the input coupler showing the E probe coupling to a short resonant length of coaxial waveguide	276

9.9	Convergence behaviour of the cavity	277
9.10	Diagrams of the fields in the input cavity and coupler for the operating mode	278
9.11	Plot of E_z around a circle passing through the beam centres. Showing the asymmetry caused by a single coupler.	279
10.1	Coaxial cavity schematic (no beam pipes)	282
10.2	“Solid” vacuum representation of the couplers with the ports marked	287
A.1	The original plots which allowed selection of the optimal cavity (w.r.t. R/Q). These plots require a large number of simulations and can be hard to read.	291
A.2	Some simulation ranges can yield poorly defined R/Q curves using the brute force method (with 1.3 GHz contour). This is a clear illustration that a new optimisation method is needed.	292
A.3	Frequency stability around 1.3 GHz of a range of simulations controlled by a polynomial approximation of 1.3 GHz cavities	293
A.4	Plot of R/Q against the independent variable h showing a further polynomial approximation representing the R/Q vs h . Allowing the R/Q and cavity size to be calculated from one independent variable.	294

List of Tables

1.1	Original CLIC specification (now out of date) [28]	5
1.2	Main klystron parameters	40
1.3	Original CLIC specification (now out of date)[28]	45
1.4	Improved TESLA MBK parameters (values in brackets are original values)[28, 29]	46
2.1	Summary of fixed klystron/cavity parameters	49
2.2	Re-entrant cavity: Dimensions (see table. 2.1)	57
2.3	Re-entrant cavity: Optimal parameters and simulation results	58
2.4	Recessed re-entrant cavity: Dimensions (see table. 2.1)	59
2.5	Recessed re-entrant cavity: Optimal parameters and simulation results	61
2.6	Fundamental mode coaxial: Cavity dimensions (see table. 2.1)	63
2.7	Fundamental mode coaxial: Optimal parameters and simulation results	64
2.8	HOM coaxial cavity: Cavity dimensions (see table. 2.1)	65
2.9	HOM coaxial cavity: Optimal parameters and simulation results	67
2.10	Whispering gallery cavity: Cavity dimensions (see table. 2.1)	68
2.11	Whispering gallery cavity: Optimal parameters and simulation results	69
2.12	Initial investigation conclusions	70
2.13	Preliminary klystron parameter problems	71
2.14	Summary of preliminary fixed klystron/cavity parameters	71

2.15	Summary of changes in the revised parameters	76
2.16	Initial investigation recessed re-entrant cavity results (see table. 2.3)	78
2.17	Initial investigation fundamental mode coaxial cavity results . . .	82
2.18	Simulation range for coaxial 2 nd harmonic cavity	86
2.19	2 nd harmonic cavity possible properties	87
2.20	Revised parameter cavity comparison	88
3.1	Modes of interest in the dipole shifted cavity	101
4.1	The boundary conditions of <i>TM</i> modes	111
4.2	Test cavity dimensions (mm)	122
4.3	Comparison between frequencies (in GHz) from MAFIA and MoM Model for azimuthal variation (<i>i</i>) = 0	123
4.4	Comparison between frequencies (in GHz) from MAFIA and MoM Model for azimuthal variation (<i>i</i>) = 1 The modes highlighted in bold are potentially mixing with each other in pairs.	123
4.5	The new boundary conditions compared to the <i>TM</i> equivalents .	126
4.6	Test cavity dimensions (mm)	131
4.7	Full table of results from <i>TE</i> and <i>TM</i> basis function simulations for azimuthal index (<i>m</i>) 1	131
4.8	The boundary conditions of <i>TM</i> and <i>TE</i> modes (both are required)	132
4.9	Comparison between frequencies (in GHz) for Azimuthal Varia- tion (<i>i</i>) = 0 from MAFIA and MoM Model with pure <i>TM</i> errors from table 4.3	139
4.10	Comparison between frequencies (in GHz) for Azimuthal Varia- tion (<i>i</i>) = 1 from MAFIA and MoM Model with pure <i>TM</i> errors from table 4.4	140
5.1	Properties and settings used for stability analysis	148
6.1	High R/Q cavity gap voltages	162

6.2	High R/Q revised klystron parameters cavity gap voltages	164
8.1	Optimisation space definition	207
8.2	Elliptical problem properties	230
8.3	Zitzler T_6 problem properties	231
8.4	Initial calibration parameters	233
8.5	Single objective sweep of σ_x simulation setup [16]	234
8.6	Search strength σ_x calibration setup	235
8.7	Updated calibration parameters	236
8.8	Mutation strength σ_μ calibration setup	236
8.9	Updated calibration parameters	237
8.10	Number of parents p calibration setup	238
8.11	Updated calibration parameters	238
8.12	Population calibration setup	239
8.13	Updated calibration parameters	240
8.14	Sweep of s simulation setup[16]	240
8.15	Calibrated parameters	243
8.16	Optimisation space definition	243
8.17	Pareto Search Pressure Focusing settings	244
8.18	TH2178 klystron specification	245
8.19	TH2178 AJDisk simulation results	245
8.20	TH2178 AJDisk post calibration simulation results	247
8.21	TH2178 initialisation range	248
8.22	Improved TESLA klystron specification	255
8.23	Improved TESLA klystron hand optimised interaction structure (see chap. 2)	255
8.24	Improved TESLA AJDisk simulation results	256
8.25	Improved TESLA AJDisk post calibration simulation results	257
8.26	Comparison of gap voltages between AJDisk and Klys2D	258
8.27	Initialisation range	259

9.1	Output coupler specification	269
9.2	Coaxial mode cut-off frequencies	270
9.3	Operating and dipole mode properties (Q_e are for cavities with 10 posts)	273
9.4	Input coupler specification	275
9.5	Operating and dipole mode properties	278
10.1	Improved TESLA MBK parameters (values in brackets are origi- nal values)[28]	281
10.2	Revised parameter cavity comparison	283
10.3	Improved TESLA klystron interaction structure	286
A.1	Simulation performance comparison	294
B.1	Comparison between frequencies (in GHz) from MAFIA and MoM model for azimuthal variation (i) = 2	297
B.2	Comparison between frequencies (in GHz) from MAFIA and MoM model for azimuthal bariation (i) = 2	298

List of Algorithms

7.1	Tournament selection algorithm [14]	176
7.2	Truncation selection algorithm [14]	177
7.3	Linear ranking selection algorithm [14]	178
7.4	Exponential ranking selection algorithm [14]	180
7.5	Basic PCX[15]	185
7.6	MrCX (bold text denotes alteration from PCX)[17]	187
7.7	Minimal Generation Gap (MGG) model[15]	189
7.8	Generalised Generation Gap (G3) model [15]	190
7.9	Basic VEGA algorithm neglecting wealth distribution and cross breeding [19]	196
7.10	NSGA algorithm [19]	198
7.11	NSGA II algorithm [20]	200
7.12	SPEA algorithm [22]	202
7.13	SPEA2 algorithm [23]	205
8.1	Spartan algorithm[16] (see fig. 8.6)	215
8.2	The rejection population management algorithm	217
8.3	The significance population management algorithm	218

Chapter 1

Literature Review

1.1 Introduction

There is general agreement among physicists that the Standard Model is the best description available of the fundamental nature of matter. However there are growing indications that the Standard Model is not able to explain all experimental observations. The problems can be grouped into Mass, the Unification of forces and Flavour (the multitude of types of quarks and leptons). Of particular interest is the origin of mass and whether this is due to the Higgs boson. It is hoped that particles colliding in the TeV range of energies will provide some useful observations, on this and other physics.[30]

The current highest energy particle accelerators are hadron colliders, more specifically proton-anti-proton. LHC¹ with a centre of mass energy at collision of 14 TeV and the Tevatron with 2 TeV are current operational colliders in this energy range, and it is hoped that these will discover the Higgs boson. If indeed the Higgs is detected, LHC does not have the precision to determine whether what has been discovered is in fact the predicted Higgs, or some other Higgs-like particle. What will be needed is a more precise machine capable of making further measurements.[31]

It is expected that many of these measurements could be made using a

¹Large Hadron Collider

lepton-anti-lepton collider which would allow the elementary particles to collide directly (a proton being a composite of quarks).[30] The well known properties of leptons, in respect to the standard model, and ease of polarisation provide constraints which increase the precision of any measurements taken.[31]

The intrinsic technical problems involved in building a muon-anti-muon ($\mu^+ - \mu^-$) collider, such as the intense proton source and other issues such as cooling, rule out such a collider at this time. For these reasons it is widely accepted that the next particle accelerator should be a TeV scale electron-positron ($e^+ - e^-$) linear collider.[30] Such a collider must be linear to avoid the loss of energy via the emission of synchrotron radiation, this is much more severe for electrons than for protons.

There are two current designs for such a machine; International Linear Collider (ILC), an L-band (1-2 GHz) superconducting linear accelerator; and the Compact Linear Collider (CLIC), an X band (8-12 GHz) normal conducting linear accelerator.[28]

ILC uses klystrons to directly power the cavities which accelerate the beam. This leads to a complicated, and expensive, beam tunnel design with the RF (Radio Frequency) sources placed along side the beam tunnel (see fig. 1.1a); this is further complicated by the use of super conducting cavities.[28]

The Compact Linear Collider (CLIC) study is developing the technology for 0.2 TeV - 3 TeV electron-positron collider based on the Two Beam Acceleration (TBA) principle. It is being designed with a high centre of mass energy to complement the LHC and continue investigating its findings. In contrast to the ILC the cavities are not directly powered by the klystrons. The RF sources are instead used to bunch a "drive beam" from which the energy is extracted to accelerate the electrons.[30] Figure 1.2 shows the proposed layout of CLIC.

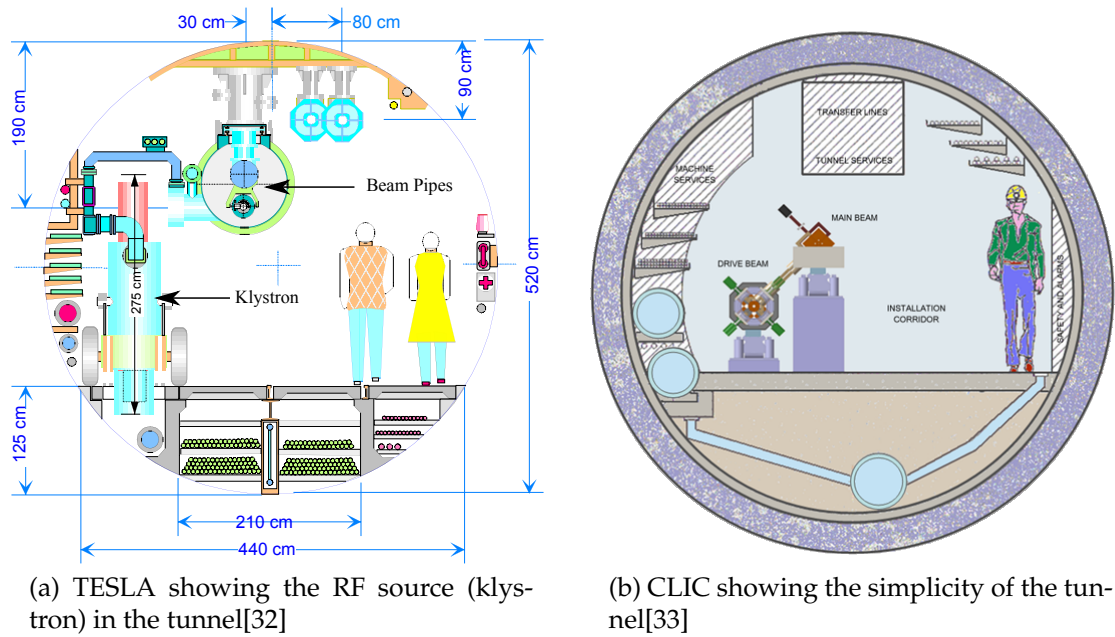


Figure 1.1: Tunnel Configuration of a direct acceleration (TESLA) and TBA (CLIC) accelerator showing the relative complexities

The operation of this is similar to that of a transformer. A low-energy, high-current drive beam travels parallel to the main beam, which is accelerated by transfer of energy from the drive beam via decelerating RF structures called PETS (power extraction and transfer structures).[33] To achieve the high collision energy cost effectively, CLIC must accelerate the electrons using an accelerating gradient of 100 MV/m, this is only feasible using high frequency resonant structures.[33]

The pulsed electron drive beam is accelerated to 2.38 GeV over 1km in the drive beam accelerator. This is powered directly by 326 klystrons operating at 33 MW. The resulting beam is bunched at 1 GHz, using combiner rings and a delay loop, this beam is converted into a 12 GHz 68 MW beam. This is distributed to power the accelerating structures of the main linac via the PETs.[33] This approach has the advantage of having no modulators or klystrons alongside the linac (they are placed in a separate facility) enabling the main linac to be placed in a single small, easily extendable tunnel (see fig. 1.1b).[1]

The TBA principle also allows the designers to avoid using X-Band klystrons, instead simpler L-Band klystrons can be used as the drive beam's fre-

quency is multiplied by the use of combiner rings.

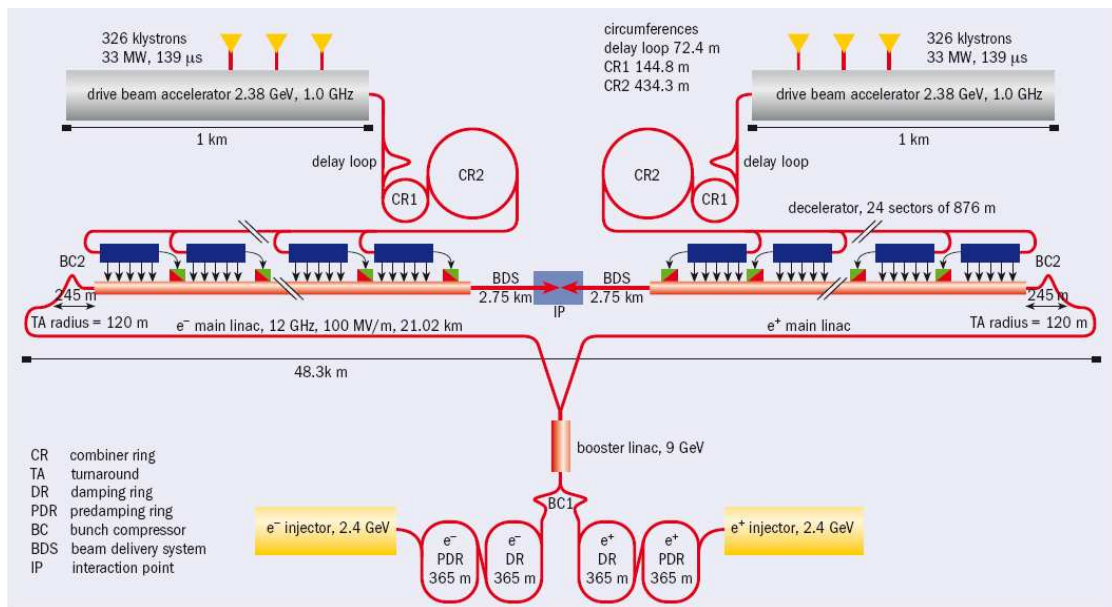


Figure 1.2: Overall layout of the CLIC complex at 3 TeV[1]

The current specification of CLIC requires a high power drive beam accelerated using RF sources at 1.3 GHz. A number of devices are often used as accelerator RF sources, the tetrode, the IOT (Inductive Output Tube) and the klystron. In the case of all of these devices the modulating signal is significantly smaller than the resultant signal, thus they operate as amplifiers.[34]

A tetrode is a refinement of the triode, where a grid modulates the electron flow between two plates with a potential applied across them. As the electron flow across the plates is modulated so is the voltage. However this device cannot reach the frequencies required due to transit time limitations.[34]

An IOT uses a gridded gun to modulate an electron beam and an RF cavity to decelerate the beam and extract the energy from it. An IOT can reach high operating efficiency but the gridded gun and the beam modulation limits the available power.[34]

A klystron uses a cavity to modulate the beam before using another to extract the power. Klystrons should be able to supply both the power and efficiency required at the correct frequency. The beam bunching (as opposed to

beam modulation) allows the current in each bunch to be almost all double the unmodulated beam current.[34]

Not all of these devices are suitable; the specification (see table. 1.1) however is beyond the current state of the art and further research and development is required to achieve the required efficiency and power output.

Table 1.1: Original CLIC specification (now out of date) [28]

Parameter	Value	Units
Operating Frequency	937.5	MHz
RF Pulse Width	100	μs
Pulse Repetition Frequency	100	Hz
Klystron Peak RF Power	50	MW
Klystron Average Power	500	kW
Efficiency	≥ 65	%
Gain at Saturation	≥ 43	dB
Beam Voltage	< 220	kV

1.2 Klystron Operation

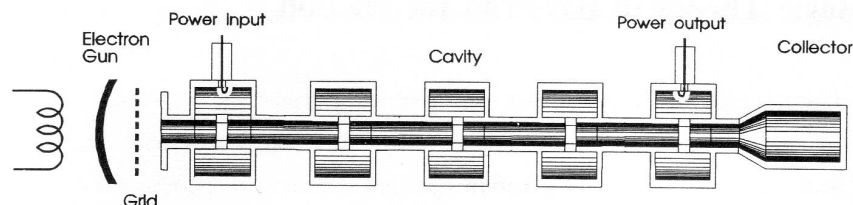


Figure 1.3: Single beam klystron schematic[2]

The main type of high power amplifier used as an RF source for high energy particle accelerators is the klystron, for instance a 1300 MHz 10 MW source designed by Thales.[34] The basic idea behind the operation of a klystron is deceptively simple, an input cavity bunches an electron beam which when it arrives at the output cavity excites an electric field much larger than the field which caused the bunching, thus acting as an amplifier.

In more complex terms a small RF signal is coupled to an input cavity with a similar resonant frequency to its fundamental mode ($TM_{0,1,0}$). As the electron beam passes through this cavity, and depending on the phase of the electric field, it is either accelerated or decelerated. This results in a loosely velocity modulated beam which, if space charge was to be ignored, would converge into tight bunches before diverging again (see fig. 1.4a).[2]

With the introduction of space charge, the electrons repel each other as they converge and the bunches are limited in how tight they can become (see fig. 1.4b). At this point any overtaking is unlikely, apart from at exceptionally high drive levels (high input power). Very crudely the cavities should be placed at points where the bunching is greatest. This is certainly true for the output cavity.

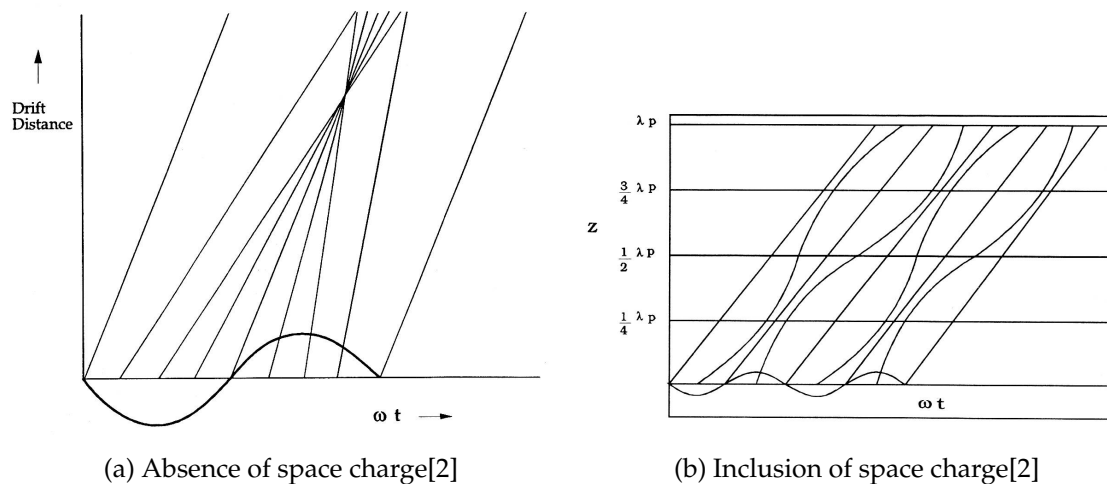


Figure 1.4: Applegate diagram showing convergence of electrons in time for a single RF phase[2]

This modulated electron beam travels down the drift tubes² to the output cavity which acts as a decelerating structure, converting the kinetic energy from the beam into RF energy in the cavity. This is achieved by using the high charge density in the bunches to excite RF currents in the cavity. This is then coupled out using an appropriate transmission line.[2]

The tightness of the bunching is critical to the efficiency of the tube and can be greatly increased with the inclusion of intermediate idler cavities. These

²or beam pipes

extra cavities can, if the resonant frequencies and positions are chosen with care, not only increase the efficiency, but also the bandwidth.[2]

Due to the way the bunches are formed the beam has a high harmonic content, particularly so in a broadband klystron. The cavity frequencies (and their higher mode frequencies) must not coincide with the higher harmonics of the beam, as this will cause spurious signals to be amplified.[2]

A second harmonic cavity is often included as the third cavity to increase efficiency. This cavity causes electrons left in the "anti bunch" to be collected into the bunch.[34]

1.3 Components of a Klystron

1.3.1 Electron Gun

The electron gun is responsible for the generation of the required electron beam. The beam must be well formed as it leaves the gun region and enters the focusing field.[2]

An electron gun operates by emitting electrons from a source, the cathode, and accelerating them using a voltage applied between the cathode and anode (see fig. 1.5). The shape of the anode and a focusing electrode form the beam to the desired size.[2]

The key defining parameters of the gun are the beam voltage V_b (which defines the electron velocity), the beam current I_{b0} and the beam radius b . [2]

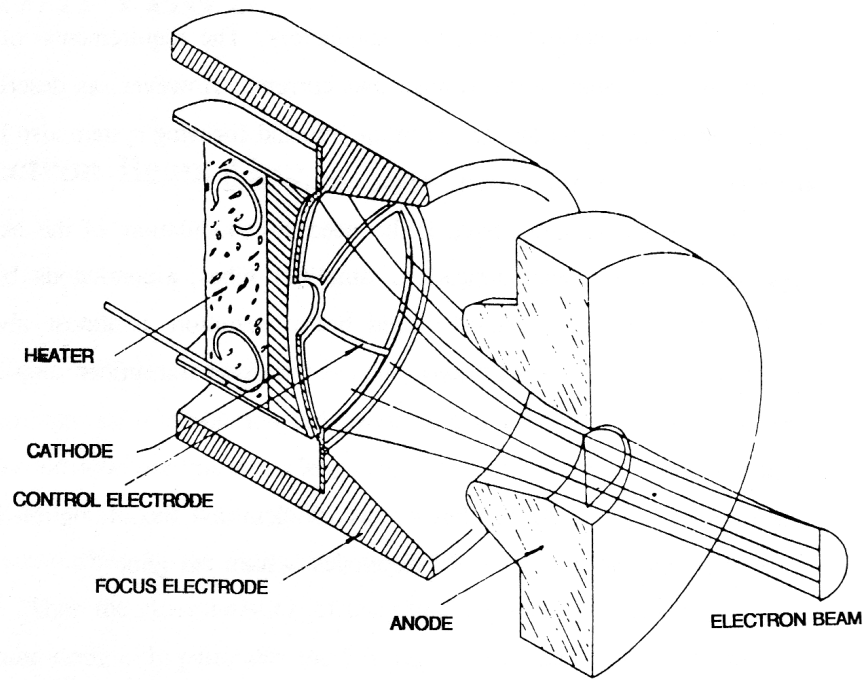


Figure 1.5: Diagram showing the components of an electron gun. The control electrode is not used on high current cathodes.[2]

In many systems, such as broadcast of communication, continuous RF power is required. For others, such as radar, pulsed power is advantageous. This can be achieved by modulating the RF, or modulating the beam itself. Klystrons for scientific use are often pulsed, for efficiency and due to the design of the accelerators which they power. Pulsing of the beam can be achieved by modulating the acceleration voltage, or by the use of a control electrode close to the cathode surface to suppress emission. In high power scientific klystrons such an electrode would be quickly destroyed, so either the high voltage must be modulated or the beam left on and the RF source modulated.[2]

Cathode loading is a measure of how much current is produced per area of the cathode. This must be kept within strict limits to avoid shortening the life of the cathode surface.

$$j_c = \frac{I_{b0}}{A} \quad (1.1)$$

where I_{b0} = the beam current and A = the area of the cathode.

The area convergence of the gun $conv$ is defined by:

$$conv = \frac{r_c^2}{b^2} \quad (1.2)$$

This is a measure of how much the gun focusing fields compress the beam after it leaves the cathode. Poor selection of this value leads to complex or poor beam focusing, although it is often mandated to some extent by beam size and cathode loading limitations.[2]

1.3.1.1 Thermionic Emission

The mechanism by which the cathode produces electrons is thermionic emission or thermal electron emission, which is the emission of electrons induced by heat. This occurs when the energy given to the electrons by a heat source exceeds the forces restraining them, causing them to be emitted from the surface. The rate of electron emission from the cathode surface is described by Richardson-Dushman equation:

$$J = 120T^2 \exp\left(-\frac{e\phi}{kT}\right) \quad (1.3)$$

where $J =$ current density in $A \cdot cm^2$, $T =$ temperature in Kelvin, k is Boltzmann's Constant, e is the charge on an electron, $\phi =$ is the work function of the cathode surface in electron volts.[2]

The temperature is the temperature of the cathode surface and for suitable emission using existing cathode technology, this is kept at around 1200 K. The work function ϕ of a material is the minimum energy required to remove an electron from a material. Materials which have a high ϕ and remain solid at 1200 K are limited.[2]

1.3.1.2 Cathode Chemistry

The chemistry of the cathode comes from a desire to have as low a work function as possible; to achieve as high emission as possible; at as low a temperature as possible as reliably as possible.

In early cathodes tungsten was used with a work function of 4.5 eV. At 1200 K this would give a current density of only $1.5 \cdot 10^{-11} \text{ A}\cdot\text{cm}^{-2}$, needing 2526 K to achieve just $1 \text{ A}\cdot\text{cm}^{-2}$. [2]

It was found that a reduced work function could be achieved using an oxide of alkaline earth metals, commonly barium or strontium. This is applied in carbonate form as a coating for a nickel cathode, which is then reduced to the oxide during processing. This is referred to as an oxide cathode and can provide current densities of $0.3 - 1 \text{ A}\cdot\text{cm}^{-2}$ continuously. In operation the thin coating releases barium which evaporates. This coating is the only supply of barium which severely limits the life of the cathode, especially at high current density. Ion bombardment can strip the coating and it can also be easily poisoned by gasses during construction and operation. [2]

The L-Type cathode attempts to solve these problems by releasing the barium from a reservoir of $(\text{BaSr})\text{CO}_3$ behind a porous tungsten cathode. When the reservoir is heated the carbonate breaks down into barium oxide and, by reaction with the tungsten, free barium is released to form a layer on the cathode surface. The work function for such a cathode is 1.5 eV, this allows high theoretical emission with reliable operation around $4 \text{ A}\cdot\text{cm}^{-2}$. A major problem with this design is difficulty in sealing the reservoir. [2]

The B-Type or dispenser cathode removes the problems associated with sealing the reservoir by using the tungsten plug itself as the reservoir. Barium oxide is stored directly mixed with oxides of aluminium and calcium. This cathode yields a work function of 2.05 eV and a current density of $4 \text{ A}\cdot\text{cm}^{-2}$. [2]

B-Type cathodes have been adopted instead of oxide cathodes since the late 1980's. In an investigation of the advantages of using a B-type cathode compa-

red to an oxide coated cathode found that gun failure due to voltage breakdown is reduced from 70% to 3%.[35, 36]

The M-Type cathode is a development of the B-type dispenser cathode where the cathode surface is coated with osmium, rhenium, ruthenium or iridium. The work function is lowered by 0.1 eV and allows the cathode to be operated at 1200 K as opposed to 1340 K for a B-Type. This type of cathode is widely used in high power klystrons.[2]

The life of the cathode is also improved, but ion bombardment can remove the osmium layer causing it to revert to B-Type with its associated higher operating temperatures.[2] M-type cathodes also suppress excess barium evaporation and, although easy to poison by gasses and metallic vapours, can be repaired by ageing.[37]

1.3.1.3 Cathode Lifetime

Cathode loading is always a problem in high power electron devices. Due to the limits on voltage, beam power is often increased by increasing the current. The radius of the cathode (r_c), and so the gun geometry, is defined by the cathode current density (j_c).[2]

$$r_c = \sqrt{\frac{I_{b0}}{\pi j_c}} \quad (1.4)$$

where r_c =cathode radius; I_{b0} =beam current; j_c =cathode current density

A cathode with a high current density (obtained by increasing the cathode temperature) has a dramatically reduced reliability and lifetime due to the more rapid depletion of the (BaSr)CO₃ reservoir. A high convergence gun can be used to spread the current over a larger area, but this significantly complicates electron gun design.[2]

The reliability and lifetime of a cathode is one of the most important properties from a commercial point of view. Although there are several potentially superior cathode types, the M-Type is often used due to its proven reliability.[2]

If the cathode loading is 2 A/cm^2 there is the potential for 145,000 hours of operation as opposed to 6 A/cm^2 which has the potential for only 30,000 hours (see fig. 1.6).[3, 38]

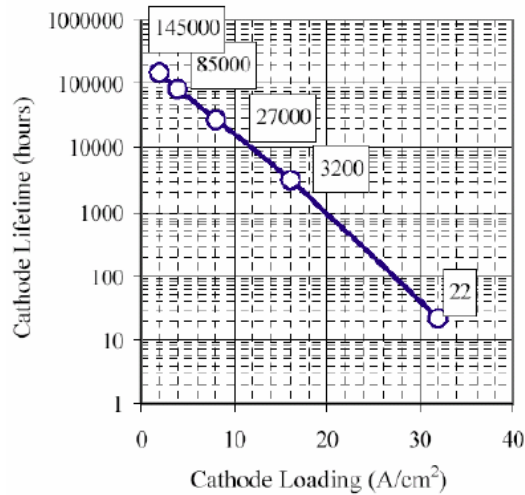


Figure 1.6: Cathode lifetime with respect to cathode loading[3]

1.3.1.4 Cathode Operating Regimes

The current drawn from the cathode is not only related to its emission properties as might be expected, there are in fact two distinct regions where different conditions dominate.[2]

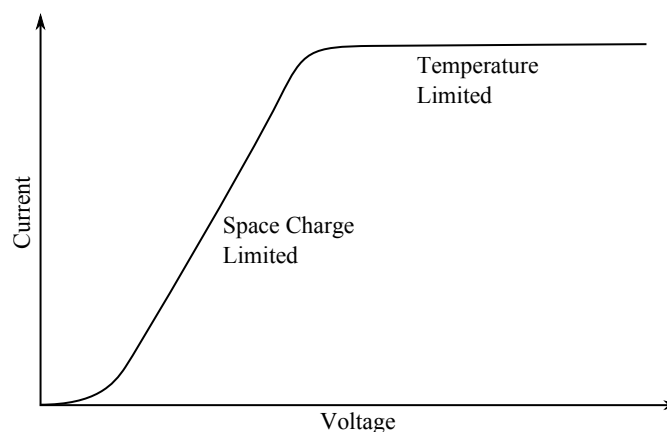


Figure 1.7: Voltage vs. current for an electron gun showing the space charge and temperature limited regions

In the absence of an electron beam, the electric field in a planar diode va-

ries linearly from the cathode to the anode. The relationship between current density (J) and voltage (V) for an planar diode with infinite plates separated by distance d is described by the Child-Langmuir three-halves power law.[2]

$$J = \frac{4\epsilon_0}{9d^2} \sqrt{2\eta} V^{3/2} \quad (1.5)$$

where $\eta =$ charge to mass ratio of an electron.

With the introduction of electrons this field is lowered around the cathode and the gradient there is reduced due to space charge. The system will then reach a state where no electrons on the cathode surface are attracted to the anode, and the electrons which reach the anode have escaped from the cloud surrounding the cathode. This occurs for lower values of V and is called the space charge limited region where $I_{b0} \propto V_b^{3/2}$. [2]

Klystron cathodes are predominatly run in the space charge limited region to make them more predictable. The current only depends on the cathode geometry, rather than on the temperature or work function (which changes over time). If the high electric field gradients were experienced by the cathode surface the thin coatings could be pulled off.[2]

At higher values of V_b the emission falls short of that predicted by space charge limited operation and emission becomes dependant on thermal emission, this is known as the temperature limited region.[2]

1.3.1.5 Perveance

The constant of proportionality of $I_{b0} \propto V_b^{3/2}$ in the space charge limited region is known as the perveance K (often referred to as the micro-perveance $\mu K = K \cdot 10^6$) and is a result of the Child-Langmuir law (see eq. 1.5).[2]

$$K = \frac{I_{b0}}{V_b^{3/2}} \quad (1.6)$$

For a given power the beam current increases and the beam voltage de-

creases with increasing perveance. This causes a corresponding drop in efficiency that can be explained by the increase in space charge and the bunches becoming harder to form and maintain due to the increased repulsive forces. Since high efficiency operation is only possible with a highly modulated beam, this situation is to be avoided. Consequently in order to create high efficiency klystrons a low perveance is required.[2]

Ideally (see eq. 1.6) a low current and a very high voltage would be used to obtain the required power while maintaining the low K (this approach is used for many super power klystrons). However there are practical limits on the value of V_b , such as voltage breakdown (see sec. 1.3.5.1) which also limits the maximum current. With a single beam klystron low perveance, high power operation is difficult (requires very high voltages) which leads to the idea of multiple beam klystrons (see sec. 1.4). This allows a set of low perveance medium power beams to be combined efficiently into a high power, high efficiency device. Effectively the space charge forces would be spread out to lessen their effects.

The relationship between efficiency and perveance can be shown by the empirical formula (see eq. 1.7) which has been found to be a good fit for the data obtained from a variety of different klystrons. It suggests efficiencies of 80% are theoretically possible for low perveance beams.[3]

$$\eta_e = 0.78 - 0.16 \cdot \mu K \quad (1.7)$$

This is only one interpretation of the results and many such relationships exist.[3] It is of course imperfect but gives a good indication of what performance can be expected.[39]

1.3.2 Beam Confinement

As the electron beam traverses the interaction structure, space charge forces will, if unopposed, act to increase the beam radius. This will cause electrons to

be intercepted by the klystron structure, lowering the efficiency and potentially causing damage to the structure. In order to prevent this from happening the beam must be confined using a suitable field, for in only the smallest of tubes can this be avoided.[2]

The beam can either be focused using an electric or magnetic field. The latter being the most common in high power klystrons. In such a focusing system there are three effects at play:[2]

- electric field, from the space-charge, expanding the beam
- a magnetic field parallel to electron velocity exerting a rotational force from the radial motion of the expanding space charge
- and then the rotational velocity in the magnetic field giving a inward force to maintain the beam size

The final force should at least balance against the expanding force from the space charge and the centrifugal 'force'. When the forces are balanced and there is no radial acceleration there is a stable beam, said to be undergoing space charge balanced flow. The minimum field required for this is called the Brillouin field, however this allows perturbations in the beam (perhaps due to modulation) to oscillate, due to the forces being unbalanced in a phenomenon known as scalloping. To overcome this, fields 2 to 3 times stronger than this are used in practice.[2]

The Brillouin field can be calculated as follows[2]:

$$B_b = \sqrt{\frac{\sqrt{2}I_{b0}}{\epsilon_0\pi b^2\eta_e^{3/2}\sqrt{V_b}}} \quad (1.8)$$

$$B_0 = B_n B_b \quad (1.9)$$

where b = radius of the beam; B_0 = focusing field; B_n = Brillouin field ratio; B_b = Brillouin field; η_e = charge to mass ratio of electron.

In confined flow focusing, the beam is focused by the magnetic field from the moment it leaves the cathode so substantial magnetic flux threads the cathode.[40, 9]

When fields stronger than the Brillouin field are used the focusing flux must thread the cathode for smooth beam entry. The flux lines are altered by the use of shaping pole pieces to follow the trajectories of the electrons, so there is no resultant focusing force within the gun region for electrons with an ideal trajectory.

1.3.2.1 Solenoid Focusing

A common way of providing the constant axial magnetic field is to use a wire³-wound solenoid. The klystron itself is made from non-magnetic materials such as copper and ceramics, which do not effect the field. The solenoid is bounded at each end by mild steel pole pieces, the mirroring effect of which approximates the field of an infinitely long solenoid.

For Brillouin flow the hole in the pole piece through which the beam enters (and magnetic flux escapes) must be as small as possible, so the field increases sharply. In a confined flow design the beam is focused by the field from the moment it leaves the cathode so the aperture in the pole piece must be designed to allow the correct field profile in the gun region.[2]

The current flowing through the solenoid windings causes significant heating in high power klystrons which must be cooled. A common failure mode for solenoids is overheating leading to short circuits.[2] In pulsed tubes the solenoids often cannot be pulsed due to their inductance, and so must be run continuously. Superconducting solenoids are another option but the clear practical implications of cooling several hundred such coils for the RF power supply of a linac make them unappealing.

Due to the increased risk of interception in a high power klystron around

³or tape

the output gap (due to high modulation and so high charge density) the field may need to be increased. This causes problems with solenoid design as a gap also has to be inserted at this location to allow the output wave-guide to leave the device.[2]

1.3.2.2 PPM Focusing

Creating the magnetic field using a wire-wound solenoid requires power and so decreases efficiency. During the design of their high power pulsed klystrons other teams have found that the solenoid power is comparable to the average beam power of the klystron due to the high fields required especially at higher frequencies.[8, 4, 7] To avoid this, a permanent magnet focusing system can be used. However there are a number of drawbacks with this solution. The permanent magnets are very bulky, especially when used with a large klystron and are also susceptible to damage at high temperatures.[2]

Periodic Permanent Magnet (PPM) focusing is a system where rings of permanent magnets of equal size with opposing magnetic fields are used to focus the beam. The magnetic field reversal does not affect the focusing as the force is still perpendicular to the beam and opposes the space charge forces.[2]

This approach is very common in TWTs but not often used for high power klystrons.[4] With low perveance (which is desirable) and high voltage the space charge wavelength becomes quite long. This allows a high ratio of beam plasma wavelength to magnetic period which is optimum for PPM focusing (see fig. 1.8).[4] PPM solves to some extent the size problem as using magnets with altering field directions is substantially smaller than a uniform magnet of the same strength.[2]

PPM is not completely ideal for klystrons for a number of reasons. The best configuration for a PPM system uses a constant period, but the klystron has different drift lengths between its cavities and placing an integer number of the magnet lengths between them constrains their spacings. The PPM system also

gives more ripple in the beam due to the non uniform field, which will lower efficiency and may cancel out the efficiency gains. Although the complex analysis of the PPM system may also rule it out for such a high power and high efficiency tube, it was suggested for the SLAC 1.5 GHz MBK[41] and experimental tubes were built with some success[42].

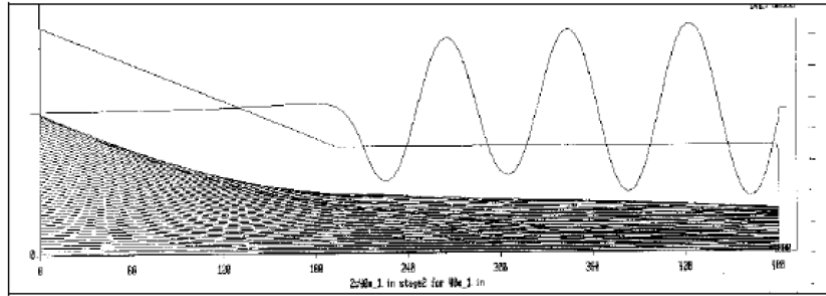


Figure 1.8: PPM focused beam [4]

1.3.3 Output

The purpose of the output structure is to decelerate the bunched electron beam and convert its stored kinetic energy to RF which can be extracted from the structure using some kind of transmission line. There are three potential structures used for extracting the energy from the beam.

A single cavity output where the cavity is tuned to the target frequency and, with correct external coupling decelerates the electrons, extracting their energy in the form of RF. [2]

A coupled cavity output circuit is similar to a single cavity circuit but there is more than one gap used. This can increase the efficiency and the bandwidth of a tube. Such a structure can operate in more than one mode. The 0 or 2π mode where the cavities are in phase, the π mode where the cavities are 180° out of phase.[2]

Taking inspiration from the Travelling Wave Tube (TWT) a travelling wave structure has been used as a deceleration structure. Instead of a cavity a slow

wave structure is used to match the velocity of the decelerating wave with that of the beam.[2]

Output circuit design is the most serious limitation to the power capability of the klystron. To avoid breakdown the field must be distributed over a long gap.[43]

1.3.4 Windows

Often power is coupled from the output cavity directly to waveguide. To avoid the inclusion of the full length of waveguide in the klystron vacuum envelope, the waveguide is sealed using RF windows. The simplest window is made from a half wavelength thick sheet of dielectric, although losses would be unacceptably high. Windows are designed for optimal matching and minimum RF losses.[2]

The RF windows are one of the most delicate components of a klystron accounting for 25% of klystron failures, and as such are critical to the life time of the tube.[3] Output windows need to be made from a low RF loss, refractory material. They are often made from high grade alumina or beryllia⁴ in cases where increased power dissipation is required. It is important to consider multipactor (see sec. 1.3.5.2) when designing output windows as it can cause high thermal stress and damage the windows.[2] Mechanical stress in the windows must be considered, sharp corners and temperature gradients due to window cooling can lead to window failure and potential vacuum loss.[2]

50 MW peak power required for CLIC means that successful design of the output windows is not trivial.[3] Even at the 150 MW extracted from a derivative of the SLAC 5045, designed for DESY, power is extracted through four windows fed by two waveguides with a peak power of 38 MW through each window. The repetition rate however was only 60 Hz reducing the average power to 2.5 MW.[44]

⁴often avoided due to its toxicity

The maximum field strength allowable on the window surface is 8 kV/mm with highly purified Al_2O_3 . [45] If during the design of the klystron it is found that this may be exceeded, a simple solution may be to increase the size of the window. This solution was employed by a team at KEK who found some tracks showing evidence of breakdown near the edge of the window in the direction of the electric field. [45]

Various tubes have been designed as an extension of the SLAC 5045, a well tried and tested tube. The approach of keeping stresses on components such as the windows below that of an existing reliable device is preferable. [44]

The Multi-megawatt Mini-windows concept spreads the power and thermal stress over a large number of windows which, coupled with the fact there is no significant electric field at the windows should lead to improved reliability. [3] There is the disadvantage that there could be reliability problems due to leaks with the large number of joints in the structure, and the structure is only applicable for certain modes.

Multipactor suppression (see sec. 1.3.5.2) and charge build up can be alleviated by treating the windows surface with TiN coating to lower secondary emission yield. [43, 45] If the RF windows were to be horizontal there is the possibility that there could be an increased risk of breakdown due to "foreign matter" building up on their surfaces. [46]

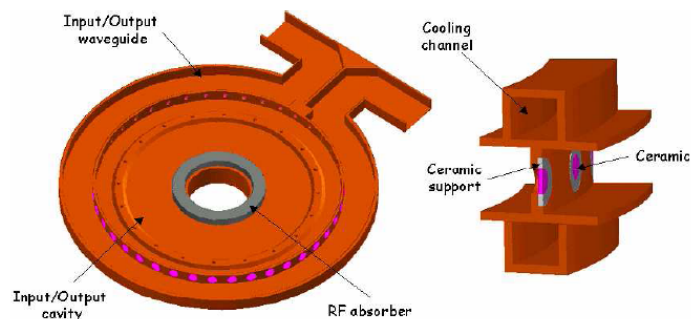


Figure 1.9: View of output cavity with mini-windows [3]

1.3.5 Limiting Factors

1.3.5.1 Voltage Breakdown

Voltage breakdown occurs when the electric field between two points exceeds the maximum breakdown electric field of the vacuum. The maximum voltage hence increases with distance. This is believed to be caused by ionisation of evaporated material from the wall, usually initiated by field emission and its associated heating.[47] This limits the voltage which can be used at the cathode in the pursuit of a low perveance (see sec. 1.3.1.5). Not only does breakdown reduce the efficiency of the tube and disrupt its operation, but it often leads to irreparable damage such as pitting.

In a klystron, beyond the gun, the most likely region to experience breakdown is in the output gap. The maximum permissible safe electric field in the output gap, as used by Thales, is 10 kV/mm.

There is an empirical formula for reliable operation, with respect to breakdown, applicable to generic RF structures called the Kilpatric Limit.[48]

$$f = 1.6E^2 e^{-8.5/E} \quad (1.10)$$

where E is the electric field in kV/mm below which reliable operation can be expected and f is frequency in MHz. The Kilpatric Limit was derived from data produced prior to the the 1950s, short pulses and improvements in equipment, vacuum pumps for instance, allow this limit to be exceeded.[49]

1.3.5.2 Multipactor

Multipactor is the name given to the phenomenon where secondary electron emission resonates with the electric or magnetic fields. It was discovered by Farnsworth in 1934 who exploited the effect in a high gain amplifier in his early television camera tubes. He also accepted that unintentional multipactor could result in damage to an electron device.[50]

Multipactor often occurs between one or two surfaces in an electron device. The emitted electron is accelerated by the electric field and strikes the tube surface causing a secondary electron to be emitted, which depending on the phase of the field, could be accelerated back to cause another secondary emission. For this to occur on two surfaces the transit time for the electron must be roughly equal to $NT/2$ (where $T =$ one RF period, $N = 1, 3, \dots$). Another type of multipactor is where the electrons are deflected back to where they came from in an integer number of wavelengths of the RF field, this is only possible under certain conditions.[50]

Multipactor causes a number of problems, the most obvious being damage to the tube or output windows caused by localised heating due to interception. For less severe multipactor a marked drop in efficiency can be seen around a certain frequency. For high efficiency and reliable operation it is clear that multipactor should be avoided as much as possible.[34]

Multipactor is not currently easily simulated so avoiding it completely is difficult. The best method of suppression is to find where the multipactor is taking place and appropriately alter the geometry of the structure, perhaps adding castellations or applying a coating.[2]

1.3.5.3 Monotron Oscillations

The original monotron was a discrete device consisting of an electron beam sent between two parallel grids. The sharp irregularities cause the beam of electrons to excite EM fields and the system begins to oscillate.[51]

Monotron oscillations occur in a cavity when the interaction gap is long enough for a beam to be both bunched and then decelerated in one cycle. This causes the energy stored in the cavity to increase unless it is damped.[52]

The start condition is when the transit angle θ is between 2π and 3π , 4π and

5π , etc. Where the transit angle is:[51]

$$\theta = \frac{\omega g}{v_0} \quad (1.11)$$

where $\omega =$ is the angular frequency of the mode; $g =$ is the gap length; $v_0 =$ the velocity of the electron as it reaches the gap.

If this condition is met in a klystron an unbunched beam can become bunched by a single cavity, and the klystron will spontaneously oscillate. The cavities must be analysed to ensure that there are no modes which allow this to happen.[53, 51]

1.3.6 Examples

MBKs have been developed for TESLA by Thales and Toshiba using two output waveguides and simple pill box type windows with a peak power of 10 MW with success.[54, 29, 55] Higher power klystrons may take a different approach, for instance SLAC XC6, before being abandoned because of gun breakdown problems, produced 90 MW with two uncoupled re-entrant cavities in tandem, each with two output windows.[43] A previous tube XC1 showed evidence of "significant beam erosion and RF breakdown" in the output cavity during an autopsy. It was found that in order to suppress this breakdown the gap would have to be much longer than is possible with single cavity without producing negative beam loading (where power is lost from the cavity into the beam).

SLAC endeavoured to solve the problem of high voltage in one of their klystrons, XC2, by changing to a dual gap coupled cavity output circuit operating in the 2π mode, which reduced the peak cavity fields by 60%.[43, 56] The 2π mode being chosen because there is no net current flow, as opposed to the π mode which has severe electric field asymmetries.[57] A 5 fold increase in power over the Toshiba E4736 was attempted in the E3746 which used a single gap output design, Toshiba also considered a multiple gap output with a $\pi/2$ mode

travelling wave structure (see fig. 1.10). Since the travelling wave smoothly decelerates the electron beam synchronous with the beam velocity the power conversion efficiency is enhanced.[58, 5]

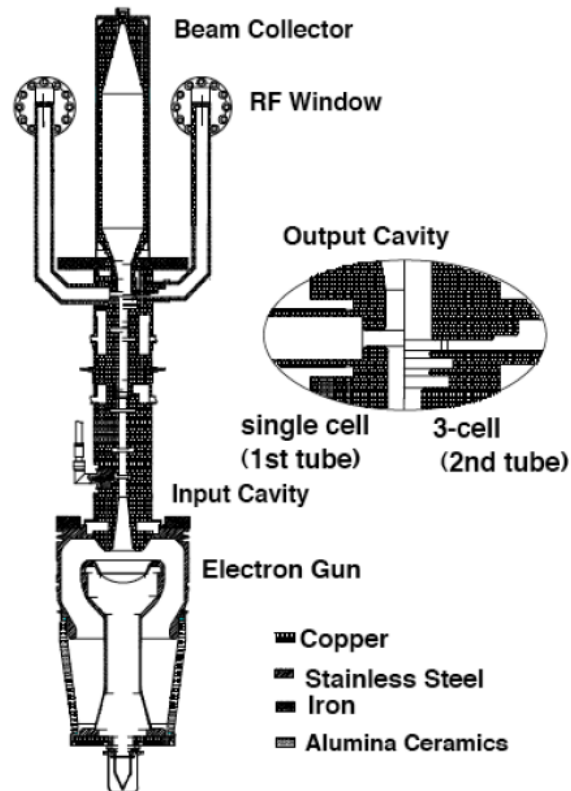


Figure 1.10: Cross sectional view of E3746 klystron[5]

A multiple beam klystron with coaxial cavities was designed as a source for JLC[6] but it was found that the R/Q in the output cavity was too low. Therefore the interaction between the cavity and the beamlets is weak. The use of a double gap output cavity improved the efficiency by increasing interaction (see fig. 1.11).[6]

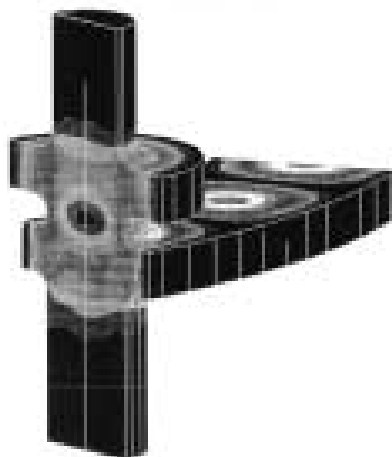


Figure 1.11: Cut through of a KEK klystron output cavity showing the double gap configuration [6]

During the design of SLAC XB72K[7] where a high efficiency low gradient output was desired a simple minded theory of a constant group phase velocity travelling wave structure was developed. The idea was to:

"let the power flow with a constant group velocity throughout the structure while evolving the beam due to [the merging] of the extracted power from [the] beam. The Q value at the output port is matched to this group velocity so that the power exits at the same speed as it flows in the structure. This smooth flow distributes the electromagnetic energy density equally."[7]

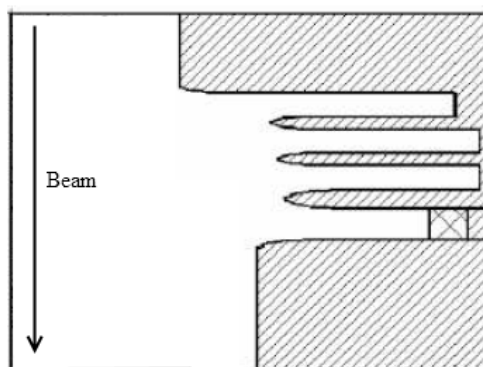


Figure 1.12: Multiple gap TW output structure in XB72K[7]

If multiple un-coupled output gaps are to be used it is preferable to extract a roughly equal amount of power from each cavity to ensure that the RF gradients are sufficiently low. If the first cavity is de-tuned slightly it has been proved experimentally that the beam travels to the second cavity (tuned to the signal frequency) well bunched. There is also a difference in phase between the output from each cavity, which although it can be compensated for, is another complication.[59]

An interesting approach to damping unwanted modes in the output circuit was employed for JLC. By tilting the output coupling slightly to break the symmetry in the output cavity they hoped to damp TE₀₁ modes.[8]

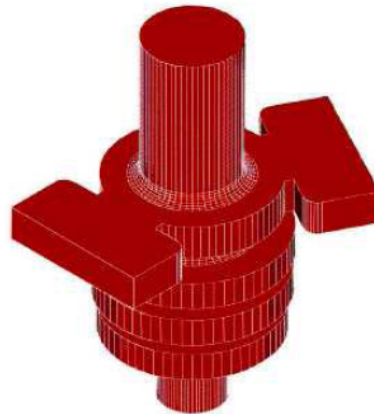


Figure 1.13: Tilted output couplers for damping of TE₀₁ modes[8]

1.3.7 Collector

The purpose of a collector is to absorb the electron beam after it exits the klystron interaction structure, without reflecting electrons backwards through the tube. Conveniently the space-charge in the beam acts to expand the beam once it is free of the magnetic field, this helps to spread the power deposited in the collector. The main consideration for a collector is to be able to dissipate the power still remaining in the beam. Smith and Philips [2] suggest that as a rule of thumb 10 MW should be the maximum time averaged power dissipated by one

square metre of collector. Although the peak beam power for a CLIC klystron is 50 MW, and so will require a $5m^2$ collector, it may be possible to pulse this to significantly reduce the average power. A collector must of course be able to dump the full beam power in the case of no modulation, as is the case when only the RF is modulated. [2]

It has been found on lower powered klystrons that the efficiency can be improved by using a depressed collector. With a conventional collector the electrons are all collected at the same potential and any energy not removed by the output cavity is dissipated as heat. With a depressed collector the electrons are slowed by a negative potential which results in an increase in efficiency as the energy is recovered and less energy dissipated as heat. However if the collector voltage is too low some of the slower electrons will be reflected and will hit the sides of the tube, a phenomenon known as back-streaming. This has the potential to cause positive feedback leading to oscillation and creates a drop in efficiency far outweighing the possible gains. Theoretically if all the electrons could be caught at the appropriate potential, the klystron would be 100% efficient. Unfortunately this is impossible in practice although this ideal can be approximated using a multi-stage depressed collector where the collector has segments at different potentials.[2]

The wide spread of energy in the spent beam of a klystron means that the probability of an electron being repelled from a depressed collector are much higher than for a Travelling Wave Tube (TWT) with its much narrower spread⁵. In principle the multi-stage depressed collector is the solution, but the complex nature of such a collector in a high power pulsed tube (mechanical and electronic) means that the improvement in efficiency might not be worth the increased and very complex R&D.[2]

One side-effect of high efficiency is the creation of slow electrons which can be reflected from the collector leading to unwanted oscillations. Care has to be

⁵around 20% for a TWT and up to 100% in a klystron[2]

taken in the design of the collector to minimise this.[29]

1.4 Multiple Beam Klystron (MBK)

An MBK is, simply put, a number of single beam klystrons in parallel. Their cavities (with the exception of the input and output, which must be common) and focusing systems can either be common or separate.

X band 75 MW single beam pulsed klystrons developed by SLAC with a voltage of 480 kV (to achieve low perveance) were at one point intended to power JLC and NLC. This high voltage (and high frequency) beam requires a high focusing field and causes problems with breakdown making it an unattractive proposition. The Multiple Beam Klystron (MBK) configuration spreads the power over a number of beams and solves these problems.[6]

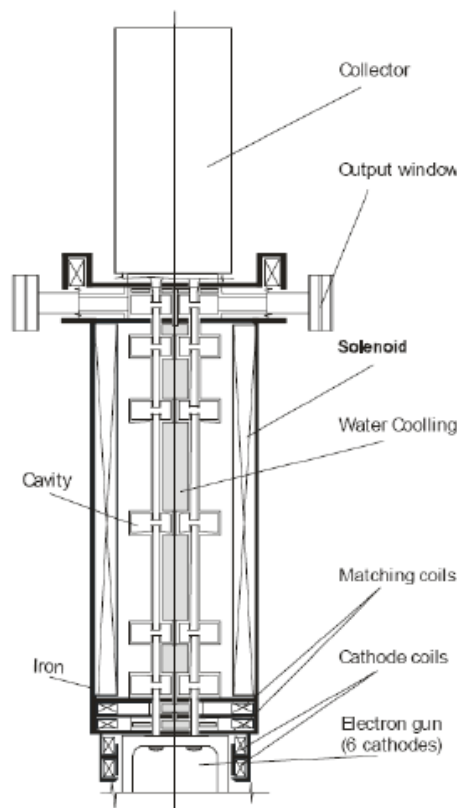


Figure 1.14: Schematic of a multiple beam klystron[9]

This configuration has a number of advantages, notably for a given output

power the gun voltage is reduced. This has the effect of reducing the perveance of each beam, allowing higher efficiencies. They are relatively compact compared to a similar single beam klystron (or group of individual single beam klystrons) and much easier to cool. [60]

The MBK configuration is suited to high efficiency operation. The need for low perveance (and so low current, see section 1.3.1.5) is satisfied as the current can be spread over a larger cathode area (several cathodes) and a suitable lifetime maintained albeit with a complicated electron gun design.

However beam focusing can be an issue as the beams are off-axis, and it is often necessary to keep the cathode loading high to allow the beams to remain as central as possible.[60]

A single common collector is likely to be the best most cost-effective design.[3] When CPI designed a 6 beam klystron for TESLA, 6 isolated collectors were used although this was mostly for diagnostic purposes and was not intended for the final models.[40]

1.4.1 MBK Focusing

Focusing is made considerably more complicated due to the fact in an MBK the axial magnetic field needs to be kept constant in order to limit the defocusing radial component, which might lead to beam interception.[29, 6]

There are a number of approaches used to optimise or tune the focusing field.[29, 6] The solenoid magnetic field could be controlled or adjusted by the use of independent sets of coils (including bucking and trim coils) to shape the field to the design value and improve beam quality, although in an MBK there must be one around each beam. In another design at SLAC the field was modified by removing the trim coil and shaping the field with a smaller diameter steel gun pole piece making the bucking coil used in this design less important.[56] However the simplest design possible is preferable. The use of a large drift tube can require a lower focusing field but can also allow spurious

oscillations due to propagating modes where the beam pipe is no longer cut off.[58]

Larionov [9] states that the majority of multiple beam klystrons use Brillouin focusing, which limits high power operation. The shielded cathodes used avoid the complexity of shaping the field in the gun region. Confined flow focusing helps achieve acceptable beam transmission for high power and high duty applications but needs magnetic flux to thread the cathode.[40]

The off axis nature of the electron guns makes it difficult to achieve the field profile required. Calabasas Creek Research corrected this field for an 8 beam singly convergent gun operating at 175 kV and 55 A using a magnetic field shaping iron (see fig. 1.15) (which also removes the need for gun coils).[10]



Figure 1.15: Magnetic field shaping iron [10]

Another solution proposed for a Toshiba MBK (E3736) is the use of two common cathode coils which are placed around the gun structure. Two magnetic lenses are then used to transport the beamlets into the focusing field which allows flexible matching with the solenoid fields. This system allows the beam diameter to be "fine tuned" to optimise the operation of the klystron.[9]

The focusing system can either be common or for individual beams. Jensen

et al. [3] suggest that each beam has its own solenoid around each drift tube. This complicates the fabrication process but may avoid problems in creating such a large volume of uniform magnetic field. A variation in the field would lead to a radial component in the overall trajectory of the beam.

PPM focusing may well lend itself to individual beam focusing (a regime suggested for a GW MBK[41] and the CLIC MBK[3]). A 150 kV 15 A beam could be focused by a PPM system using no power. By contrast if such a field was to be generated using a solenoid it would require 1 kW. This may become critical when very high overall efficiency is required.[3]

In an MBK designed at KEK for JLC a combination of PPM and coil focusing was used. The main focusing field was provided by the PPMs and was common to all the beams. A uniform field at the cathode was provided by the cathode coil which controls the beam size. They used a matching coil to optimise the transition between the gun and the PPM focused region (see fig. 1.16).[6]

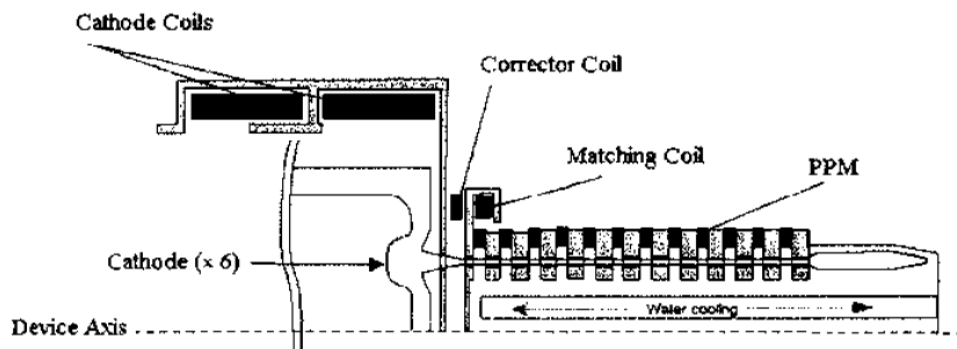


Figure 1.16: PPM focused gun[6]

The design of MBK electron guns is complicated by their off axis location. This is often particularly difficult due to the high convergence (ideally kept low) imposed by limitations on cathode current density.[43] For high power operation the MBK's lower perveance allows much lower voltages between the gun electrodes helping to avoid gun arcing.[46]

The Thales TH1801 for instance uses 7 separate M-Type dispenser cathodes and a beam forming electrode which surrounds each cathode. Special care was

taken to reduce the asymmetry of the magnetic field around each beam axis and keep the magnetic field constant (except in the compression area).[29]

1.4.2 Choice of Number of Beamlets

The choice of an MBK configuration is based on the fact that the perveance can be lowered by distributing the current. An empirical relationship suggests that a lower perveance leads to a higher efficiency, the relationship preferred by Thales is⁶:

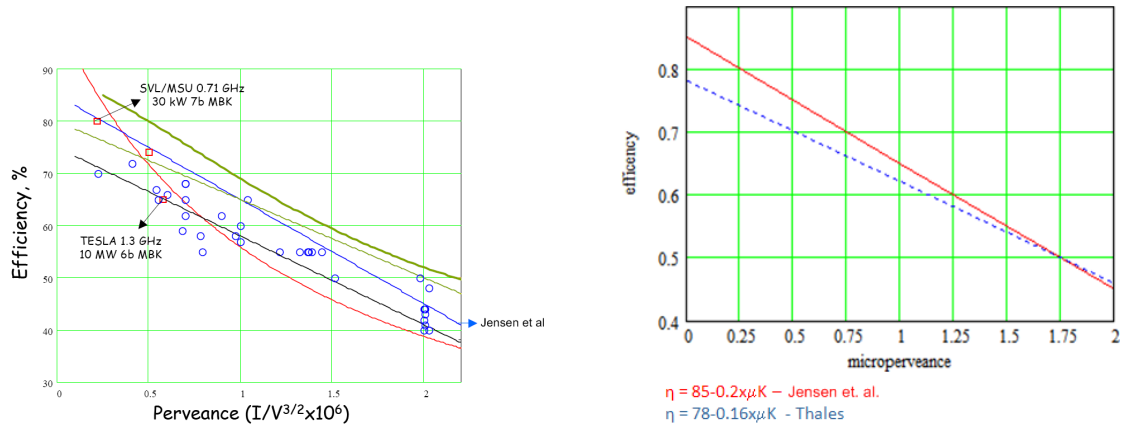
$$\eta = 0.78 - 0.16\mu K \quad (1.12)$$

This relation would suggest that an efficiency of 80% would be optimistic, however Jensen et al. suggest that the low-frequency, high-efficiency MBK data is best fitted by the slightly different empirical relationship [3].

$$\eta = 0.85 - 0.2\mu K \quad (1.13)$$

Accepting that Eq. 1.13 represents the relationship between efficiency and micro-perveance: [3] a line of efficiency vs micro-perveance can be plotted and the target efficiency of 80% selected (see fig. 1.17b). This leads to a preferred micro-perveance of $0.25 \frac{A}{V^{3/2}}$. For the power levels required this is simply not possible with a single beam klystron and without the low micro-perveance the high efficiency required is not possible.

⁶Private discussion with Thales Electron Devices Technical Staff



(a) Plot of a selection of practical klystron amplifiers showing several fits for the data including that preferred by Jensen et al. [3] (see eq. 1.13)

(b) Comparison of the relationship between efficiency and micro-perveance preferred by Thales (see eq. 1.12) and Jensen et al. [3] (see eq. 1.13)

Figure 1.17: Different ways of representing the relationship between micro-perveance and efficiency

Having selected the micro-perveance and an MBK configuration, the number of beams and their properties can be selected. For a constant power, the relationship between the number of beams and the beam voltage and current can be seen (see fig. 1.18). 20 beams is a good compromise between beam voltage, beam current (and so cathode loading) and complexity. The required voltage and currents are not beyond the state of the art and are infact reasonably conservative.

Beam Voltage (V_b)	115	kV
Beam Current (I_b)	8.15	A
Number of beams(N_b)	20	
Perveance (K)	0.209	$\mu A/V^{3/2}$

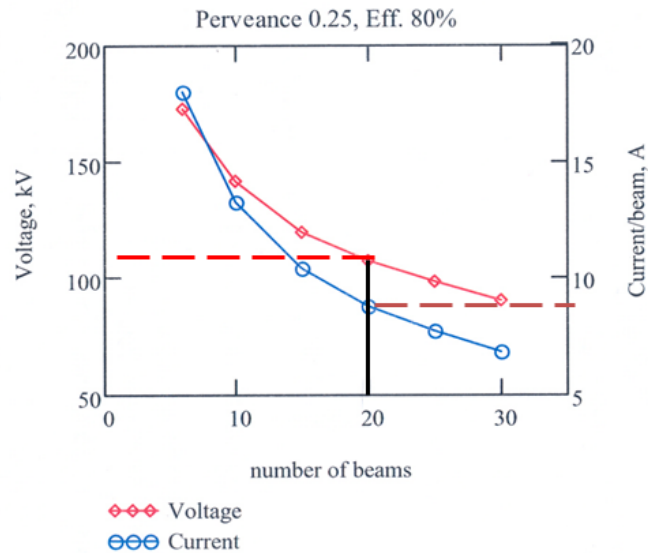


Figure 1.18: Relationship between number of beams, beam current and gun voltage at constant power and perveance. The chosen combination of number of beams, voltage and current is marked.

1.4.3 MBK Configurations

As seen in section 1.3.1.5 and more specifically equation 1.6 and 1.7 a low perveance is required for high efficiency operation. As the voltage is limited in order to maintain the power output of the device several beams are used in parallel. The literature contains a number of MBK configurations:

- $TM_{0,1,0}$ fundamental cavity (see sec. 1.4.3.1)
- $TM_{0,2,0}$ cavity (see sec. 1.4.3.2)
- $TM_{0,1,0}$ mode in a coaxial cavity(see sec. 1.4.3.3)
- $TM_{m,1,0}$ coaxial cavity (see sec. 1.4.3.5)
- $TM_{m,1,0}$ cylindrical cavity (see sec. 1.4.3.6)

1.4.3.1 $TM_{0,1,0}$ Fundamental Cavity

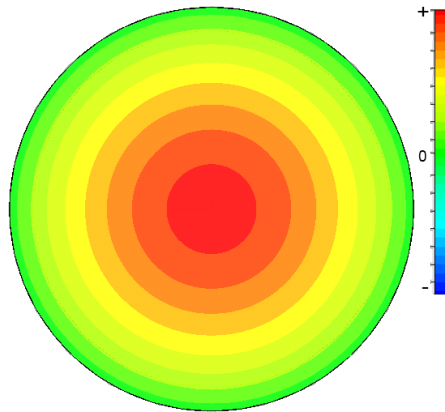


Figure 1.19: E_z intensity of the $TM_{0,1,0}$ mode in the transverse plane

This configuration uses the fundamental mode of the cavities which has a single central maximum. This places a restriction on the number of beams because they must fall inside $\approx \lambda_0/4$ for sufficient interaction.[60] This also constrains the cathode and collector diameter, causing loading and power dissipation issues. Depending on the frequency (and therefore λ) beam-beam interactions may also become an issue.[40]

Thales has already successfully designed and built a 7 beam klystron (the TH1801) using this mode for DESY with 6 common cavities and a beam perveance of $0.5 \cdot 10^{-6} \mu K$. At a peak power of 10 MW at a pulse length of 0.5 ms, tubes built to this design achieved an efficiency of around 65%. The 7th central beam interacts with the cavities' centres helping to use all the stored energy. The use of the central beam increases the voltage across the central gap.[29, 55]

1.4.3.2 $TM_{0,2,0}$ Cavity

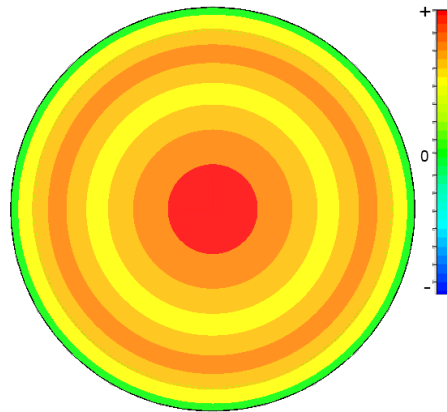


Figure 1.20: E_z intensity of the $TM_{0,2,0}$ mode in the transverse plane

The $TM_{0,2,0}$ mode has two field maxima, one in the centre and another smaller further out. CPI proposed a klystron using the $TM_{0,2,0}$ mode for TESLA because this allowed them to space the beams more widely. This allows the cathode to be larger which reduces the cathode loading, and with cleverly designed beam optics and high convergence can allow higher beam currents.[40]

Due to the larger size of the cavities their thermal dissipation is greatly improved. Using a higher order mode cavity does reduce the bandwidth of the device due to the increased mode density. Interestingly the buncher and second harmonic cavities are kept separate and work in their fundamental modes, with only input and output cavities being common.[40] The $TM_{0,2,0}$ coaxial cavities in Wright et al. [40] were found to be inherently stable with the start-oscillation current for HOMs being at least double the beam current.[3]

1.4.3.3 $TM_{0,1,0}$ Coaxial Cavity

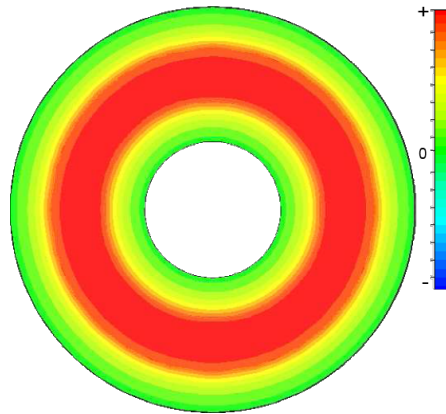


Figure 1.21: E_z intensity of the $TM_{0,2,0}$ mode in the transverse plane of coaxial cavity

Using the $TM_{0,2,0}$ or fundamental mode in an MBK with widely spaced beams both lower the R/Q of the cavity. The lack of interaction with the centre of the cavity means that the energy stored there is not exploited; ideally the energy should all be stored around the interaction area.

The Toshiba E3736 Klystron uses common ring shaped (coaxial) cavities in their fundamental modes. This allows them to separate the electron guns, increase the cathode size and so decrease cathode loading to improve lifetime. This configuration allowed a theoretical efficiency of 75% to be reached.[61, 54]

Although the Toshiba did not find any troublesome modes there is no obvious way of damping of unwanted modes in this configuration.[61, 54]

1.4.3.4 Higher Order Mode MBK

The $\lambda_0/4$ constraint can be removed from the beam locations as imposed by the shape of the $TM_{0,1,0}$ mode by using a higher order mode. If the mode is selected well it can be ensured that the energy is stored around the beam interaction areas.[60]

This allows the beams to be well spread lowering the cathode loading. However the further the beams are from the centre of the focusing field the harder

they become to focus.[60]

1.4.3.5 $TM_{m,1,0}$ Coaxial Cavity

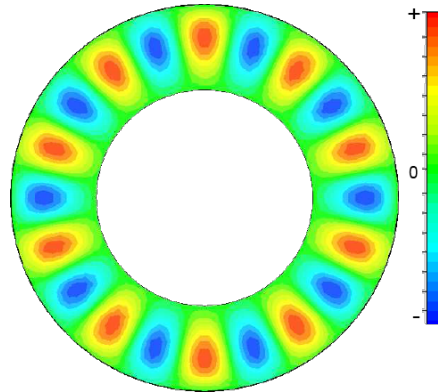


Figure 1.22: E_z intensity of the $TM_{m,1,0}$ mode in the transverse plane of coaxial cavity[3]

This is a design used by KEK [6] for their 150 MW X Band 6 Beam MBK running in the $TM_{12,1,0}$ mode. This device was expected to produce 150 MW at 55% efficiency but operated at a much higher frequency than the proposed CLIC device. The reasoning behind the choice of the coaxial mode was that the beams can be separated by appreciably more than the wavelength. Only the input and output cavities were proposed to work in this mode as the 4 intermediate cavities were designed to be independent, and work in their fundamental modes. The KEK design had problems with low beam interaction in the cavities which were solved with a double gap structure which increased efficiency by 5%.[6]

Their choice of mode ($TM_{12,1,0}$) was informed by the relatively low mode density at that point which should encourage stability. In the standing wave regime it is possible to excite the degenerate mode (90° out of phase), which would cause maxima in the B field at the beam pipes which could divert the beams into the body of the klystron. Fortunately it is found that the beam holes themselves are expected to de-tune the mode sufficiently.[6]

This mode of operation has a much denser mode spectrum compared to the

cylindrical equivalent (see sec. 1.4.3.6) which could be a significant issue as there is no simple way to damp higher order modes.[3] The number of beamlets in this configuration is equal to twice the index (m) of the operating mode.

1.4.3.6 $TM_{m,1,0}$ Cylindrical Cavity

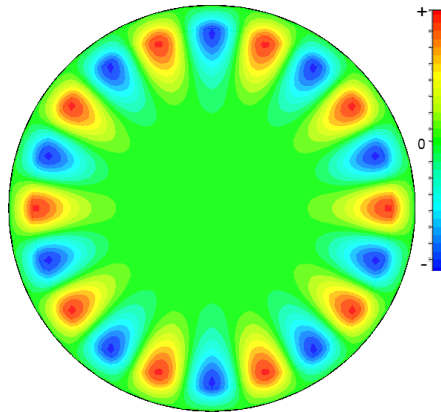


Figure 1.23: E_z intensity of the $TM_{m,1,0}$ mode in the transverse plane of coaxial cavity[3]

This cavity design has not at this time been used in a klystron, either full design or practical tube. It does however lends itself to higher order mode damping which leads to only a 10% loss in operating mode Q with 1000 times loss in undesirable modes. This design is favoured by Jensen et al. [3] for the CLIC klystrons due to the ease of mode damping.

This geometry can also support a rotating wave regime, which is in fact a superposition of two standing waves. This has the advantage of decoupling the number of beams from the operating mode and reduces the current by 25% when compared with a static wave. Conveniently if the number of beamlets chosen is odd, it discourages the excitation of nearby modes increasing the stability of the tube. However it is not clear how such a rotating mode could be reliably excited.[3]

1.5 Klystron Analysis

1.5.1 Basic Tube Properties

The main properties of a klystron determine its external properties and constraints:

Table 1.2: Main klystron parameters

Parameter	Value	Units
Operating Frequency	f	Hz
RF Pulse Width	t_{pulse}	s
Pulse Repetition Frequency	f_{pulse}	Hz
Klystron Peak RF Power	P_{out}	W
Electronic Efficiency	η_e	%
Gain at Saturation	G	dB
Beam Voltage	V_b	V

Angular frequency as always:

$$\omega = 2\pi f \quad (1.14)$$

Many high power klystrons are not CW (continuous wave) devices, the outputs being pulsed at the repetition frequency f_{pulse} . From this the duty cycle (D) can be calculated:[2]

$$D = t_{pulse} \cdot f_{pulse} \quad (1.15)$$

This leads to an average power often an order of magnitude lower than the peak output power:[2]

$$P_{av} = P_{out} \cdot D \quad (1.16)$$

When provided with an output power and a target electronic efficiency, an ideal beam input power can be calculated simply:[2]

$$P_b = \frac{P_{out}}{\eta_e} \quad (1.17)$$

This is distinct from the overall efficiency, which takes account of heater power, solenoid power and collector depression. The calculation of η_e should, strictly speaking, include the RF input power P_{in} in addition to the DC beam power, however in a high gain amplifier this is negligible.[2]

Often a beam voltage is supplied as this is limited by the high voltage power supplies available. This can define the beam current:[2]

$$I_{b0} = \frac{P_b}{V_b} \quad (1.18)$$

The beam voltage and current are subject to a further constraint in the form of perveance discussed earlier (see sec. 1.3.1.5).

As the electron velocity in high voltage tubes is mildly relativistic, this must be taken into consideration when the electron velocity is calculated from the voltage.

$$u_0 = c \cdot \sqrt{1 - \frac{1}{\left(1 + \frac{\eta V_b}{c^2}\right)^2}} \quad (1.19)$$

where η is the electrons charge to mass ratio.

The operation of a klystron can be investigated by using two techniques. Firstly the simple, intuitive ballistic model can be used. This considers the interaction between electrons and the RF electric fields in the resonant cavities. Such a model is complicated by the significant space charge concentration, causing each electron to repel to a greater or lesser extent every other. This approach is computationally intensive, although with modern computing technology and some simplification such an approach has great advantages.[2]

The second approach is to represent the electron beam as space charge waves or a plasma with a natural wavelength. The resulting equations lead to an analytical solution and is the approach used in small signal analysis.[2]

1.5.2 Small Signal

Small signal analysis makes the basic assumption that the modulation on the beam is small, allowing the non-linear terms to be ignored. As the density of electron beams used in klystrons is high, the beam can be modeled as a non-neutral plasma (or a continuum of space charge). A modulation can be imposed on this plasma, which can model the charge density and velocity without the need to model electrons, as a space charge wave.[2]

Radial debunching caused by space charge can be ignored for a magnetically focused tube as the focusing field is specifically designed to keep this to a minimum. The *longitudinal* debunching of the electrons is modelled as a natural oscillation of the plasma, rather than an effect of a collection of electron-electron interactions.[2]

This natural oscillation is called the plasma frequency, and for an infinitely long plasma it is:

$$\omega_p = \sqrt{\frac{\eta_e \rho_0}{\epsilon_0}} \quad (1.20)$$

where ρ_0 is the space charge density. If this plasma is inside a conducting cylinder (a beam pipe for example), as it is in a klystron, the frequency is reduced:[2]

$$\omega_q = R\omega_p \quad (1.21)$$

where the quantity R is the plasma frequency reduction factor. This is often approximately $\frac{1}{2}$. [2]

This plasma frequency can instruct the location of the cavities as the minimums of the beam spread can be calculated from this. Basic analysis leads to a cavity spacing of $\frac{1}{4}\lambda_q$ where λ_q is the reduced plasma wavelength (from $\lambda_q = 2\pi c / \omega_q$ for non relativistic beams) although the true optimal lies elsewhere.[2]

The gap coupling factor measures how strongly the beam couples with a

given gap length. This is due to the fact that the phase of the RF fields in the cavity changes as the beam traverses the gap. This phase change is characterised as the gap transit angle. It is possible for this to be a full cycle meaning that there has been no net energy taken from the cavity or indeed put into it. The gap coupling factor is a function of gap transit angle for a gridded gap:[2]

$$M_{\theta} = \frac{\sin(\theta/2)}{\theta/2} \quad (1.22)$$

$$\theta = \frac{\omega g}{u_0} \quad (1.23)$$

where g is the half gap length, ω is the angular frequency and u_0 is the beam velocity.

Of course if a gridded gap was used for a high power klystron the grid would be quickly destroyed. The coupling factor for a non-gridded gap where the fields can penetrate the beam tunnel is given by:[2]

$$M_{\theta} = J_0(\theta) \quad (1.24)$$

The beam filling factor also affects the coupling between the gap and the beam. A filling factor of 100% is unachievable as the beam tunnel would intercept high proportions of the beam. As the beam is always smaller than the gap it cannot interact with all of the fields. The coupling factor to represent this is M_r averaged over the beam:[2]

$$M_r = \frac{2I_0(\beta_e b)}{\beta_e b I_0(\beta_e a)} \quad (1.25)$$

The full gap coupling factor is:

$$M = M_r \cdot M_{\theta} \quad (1.26)$$

The output of a klystron using this analysis can be characterised using the

transconductance $g = \frac{I_g}{V_{g1}}$ where V_{g1} is the first harmonic of the gap voltage. From knowledge of the modulation currents (a current representing the desired modulation on the beam) the following equation can be used to calculate the trans-conductance for a 2 cavity klystron with an optimum drift length [2]:

$$g_{max} = \frac{I_{out}}{V_{g1}} = \frac{M_1 M_2 I_{b0} \omega}{2 V_b \omega_q} \frac{I'_{g2m max}}{I'_{g2m}} \quad (1.27)$$

where M_x is the gap coupling factor of gap x ; I'_{g2m} is the normalised (w.r.t. I_{b0}) modulation gap current at the 2nd gap and $I'_{g2m max}$ is the maximum. A similar equation is used to relate the RF current at the n^{th} gap due to the voltage at the $(n - 1)^{th}$ gap.

This transconductance can be combined with the impedance of the gaps (Z) and the Q_0 of the cavities to find the gain of the configuration. This gain tends to be an over estimate by 10dB when compared to the large signal gain.[2]

1.5.3 Large Signal

Large signal analysis however reverts to a ballistic model. This situation is much more complex and an analytical solution cannot be easily found. Instead iterative solutions using steady state time domain simulations of a simplified representation are used.[2]

The most basic simplification is the disk model where a series of disks of charge, with axial position and velocity, represent the beam. AJDisk is an example of such a code.[24] Radial motion is neglected and the calculation is performed in two parts, the drift between the interaction gaps and the interaction with the cavities. There are limitations to this treatment; for instance, the radial size of the beam is not constant as the bunches tend to expand due to space charge; the electric field in the gaps is not constant radially and there are quantisation errors due to the creation of the disks. It can be improved by splitting the disks into rings to take into account radial effect.[2]

A more accurate, but computationally intensive, large signal simulation method is the Particle in Cell (PIC) approach as used by Klys2D developed by Thales. This arranges the large numbers of electrons into a smaller number of macro particles which approximate the beam.[62] This allows the treatment of radial movement and avoids the kind of quantisation errors associated with disks/rings.

1.6 Klystron Specification

This study attempts to decouple the design problems specifically involved with the CLIC specification, from those more generic considerations of high efficiency operation. Klystrons following the TESLA specification have already been built and tested (for instance the TH1801[28]), whereas no klystrons have been built to meet the CLIC specification. This shows there are no insurmountable technical issues in designing a klystron to the TESLA specification. The existing klystrons built to the original TESLA specification achieve a practical efficiency of $\sim 65\%$ at 10 MW.[63] The power required by the CLIC specification is very high at 50 MW, in this thesis the TESLA specification is advanced to a higher power of 15 MW and an expected efficiency of 80%.

Table 1.3: Original CLIC specification (now out of date)[28]

Parameter	Value	Units
Operating Frequency	937.5	MHz
RF Pulse Width	100	μs
Pulse Repetition Frequency	100	Hz
Klystron Peak RF Power	50	MW
Klystron Average Power	500	kW
Efficiency	≥ 65	%
Gain at Saturation	≥ 43	dB
Beam Voltage	< 220	kV

The CLIC specification is also subject to change, in fact since the commencement of this work the operating frequency has changed from 937 MHz to 1.3 GHz, which is incidentally the operating frequency of the TESLA klystrons.

Table 1.4: Improved TESLA MBK parameters (values in brackets are original values)[28, 29]

Parameter	Value	Units
Operating Frequency	1300	MHz
RF Pulse width	1.5	msec
Pulse repetition frequency	10	Hz
Klystron peak RF power	15 (10)	MW
Klystron average power	225 (150)	kW
Efficiency	≥ 80 (65)	%
Gain saturated	47	dB
Beam Voltage	115	kV

1.7 Methodology

This study will first focus on the MBK configuration, it will then move on to more generic techniques for rapid klystron interaction structure design, optimisation and evaluation.

The literature review has shown there are a number of competing klystron configurations which are potentially appropriate for a high power high efficiency klystron (see sec. 1.4.3). After selecting a group of suitable candidate cavities, each will in turn be evaluated using Microwave Studio (MWS), a 3D Finite Difference code, over a range of dimensions to allow their comparison and the optimisation of their R/Q. From this a final cavity choice will be made to instruct the further research into the MBK.

The calculation of many modes in a resonant cavity is a time consuming process. The method of moments will be expanded to calculate modes beyond the fundamental mode to which it is currently limited. This formulation will be implemented in MathCad for comparison with MAFIA and Microwave Studio, providing a rapid and accurate analytical model for modelling any cavities which can be divided appropriately.

A manual optimisation of the interaction structure will first be performed to allow the cavity's performance to be evaluated using existing techniques. Klys2D will be used for this stage and each parameter will be optimised in turn

using parameter sweeps.

This will be followed by the development of an Evolutionary Algorithm implemented in MatLab and Java. The klystron simulations in this section will be performed by AJDisk and the high power computing requirements will be met by Condor high throughput distributed computing system. Existing klystron designs will first be optimised then the algorithm will be applied to the klystron specification used in this work.

The interaction of an MBK with the non operating modes of a cavity is not well covered in the literature and cannot be accurately simulated with any codes available to the author. A time domain model will be developed and implemented in MatLab to evaluate the significance of the dipole mode on the interaction between the beams and the cavity.

A suitable output coupler will be designed using MWS to evaluate the external Q and resonant frequency and tune it to suitable values determined by the interaction structure optimisation.

Chapter 2

Cavity Design

2.1 Introduction

The interaction structure of a klystron defines its performance. In an MBK this is highly influenced by the type of cavities used, thus the selection of a cavity type has been identified as a fundamental question for MBK design.

This decision also informs the choice of focusing system and may effect the stability of the tube. The function of this section is to compare and contrast the various cavity options with a view to making a selection. Once this decision has been made simulations of possible interaction structures are possible.

Discussions with Thales and CERN with reference to Jensen et al. [3] and sec. 1.4.3, identified 5 different cavity choices for investigation:

- Re-entrant Cavity $TM_{0,1,0}$ - as used in the TH1801 albeit with 7 beams [55]
- Recessed Re-entrant Cavity $TM_{0,1,0}$ - a novel refinement of the re-entrant cavity
- Coaxial Cavity $TM_{0,1,0}$ - as used by CPI[12]
- HOM Coaxial Cavity $TM_{10,1,0}$ - as proposed by Jensen et al. [3]
- Whispering Gallery Cavity $TM_{10,1,0}$ - as proposed by Jensen et al. [3]

While the purpose of this section is to find the maximum R/Q of the cavity for the defined constants, it is important to remember that the optimal cavity geometry will not in fact be used in the klystron. Each cavity, bar the input and output cavities, in a conventionally designed klystron is tuned slightly away from the tubes target centre frequency. It is also standard practice to alter the shape of the cavities to introduce longitudinal asymmetries or castellations in an effort to discourage multipactor.[2] With this in mind the optimal geometry need not be narrowed down very tightly for all possible optimisations performed as this is of no practical use.

In order to reduce the complexity of the problem and retain comparability between the options, only the basic cavity shapes are investigated. For instance beam nose optimisation will be reduced to create a beam ridge. Also in order to reduce the number of mesh cells the beam pipes are not simulated.

The cavities have a certain number of properties in common. To ensure comparability these are constant between the designs.

Table 2.1: Summary of fixed klystron/cavity parameters

Property	Initial
gap length (g)	10.6 mm
beam radius b (r_b)	5.74 mm
tunnel radius a (r_t)	8.21 mm
beam ridge width (r_0)	9.2 mm
cathode loading (j_c)	5.5 A/cm ²
cathode separation (cs)	3 mm
b/a	0.7
pitch circle radius (r_{pcr})	62.8 mm

These properties are constant for the initial cavity investigation. Beyond this, as noted in chapter 1, the revised parameters are used (see sec. 2.9).

2.2 Methodology

2.2.1 Modelling Fundamental Mode

MicroWave studio (MWS) is a Finite Difference Time Domain (FDTD) code.

It relies on the Yee algorithm to solve Maxwell's equations [11]:

$$\oint_{\partial A} \mathbf{E} \cdot d\mathbf{s} = - \int_A \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{A} \quad (2.1)$$

$$\oint_{\partial A} \mathbf{H} \cdot d\mathbf{s} = \int_A \left(\frac{\partial \mathbf{D}}{\partial t} + \mathbf{J} \right) \cdot d\mathbf{A} \quad (2.2)$$

The Yee cell allows these equations to be solved in turn. Eq. 2.1 can for instance be solved using a loop of electric field round one face of the electric field grid for the left hand side. Through this comes a magnetic field component which is represented by the right hand side. Next the magnetic field can be similarly calculated using the electric field component through the face of the magnetic field grid (see eq. 2.2). The eigenmode solver finds solutions to these equations for a rectangular mesh representing the cavity or other structure.[11]

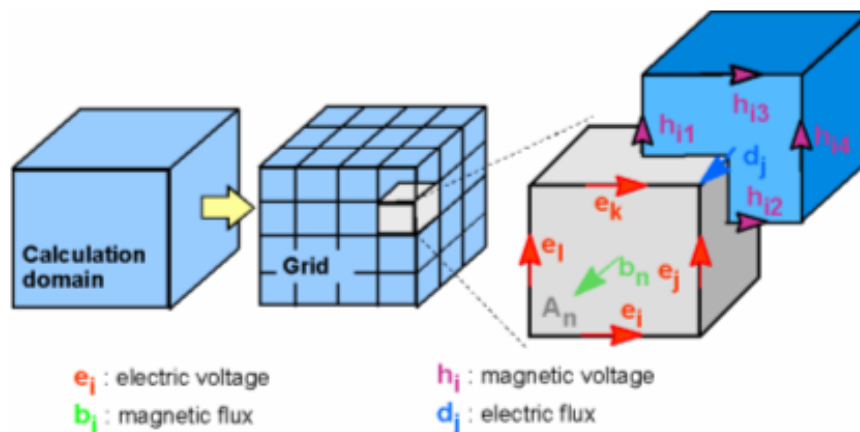


Figure 2.1: The use of offset grids to model the magnetic and electric fields[11]

The use of symmetry conditions can markedly reduce the mesh density. The fundamental mode targeted in the first three types of cavities can be found by

meshing only 1/8 of the cavity volume (see fig. 2.2). Use of symmetry conditions also increases the accuracy of the eigenmode solvers.[11]

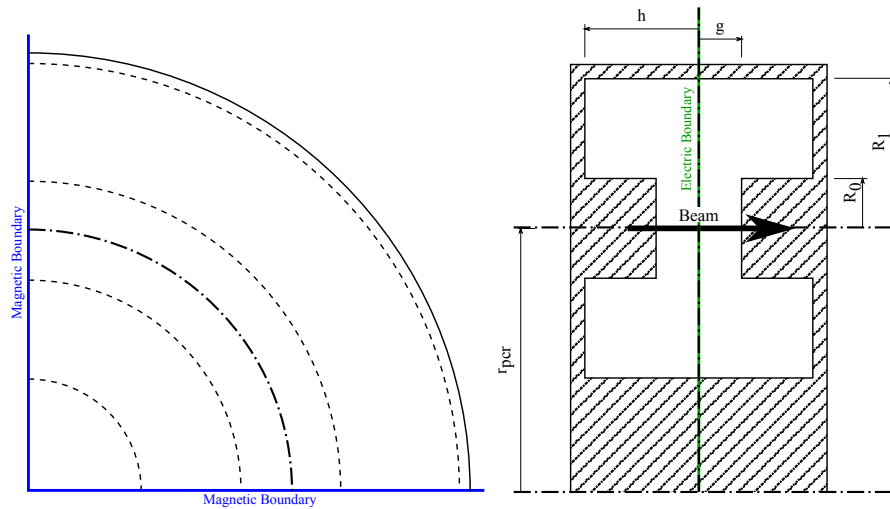


Figure 2.2: The boundary conditions used for simulations performed on the re-entrant, recessed re-entrant, coaxial fundamental cavity cavities. Green: electric symmetry plane ($E_{\perp} = 0$) at (X,Y) , Blue: magnetic symmetry plane ($H_{\perp} = 0$) at (Y,Z) & (X,Z)

From the MWS manual[11]:

“The AKS¹ method depends on an estimation of the eigenvalue of the highest mode under consideration. This estimation is chosen automatically during an iterative estimation refinement process. If many of these passes are required, it might be advantageous to choose the JDM² eigenmode solver, which is parameter free.

The solver time for the JDM eigenmode solver increases with the number of modes to calculate. Therefore, it is usually the method of choice if only a few modes are required. In many cases, the JDM solver is very robust, especially for multiple degenerated modes.”

For these reasons the JDM solver has been chosen for both the fundamental mode and HOM simulations. For the vast majority of the simulations only a single mode is of interest.

¹Advanced Krylov Subspace

²Jacobi-Davidson

A study of the convergence properties of the simulation showed that a mesh of $\sim 80,000$ mesh cells gave mesh independent frequency and a higher mesh is not required. To optimise the R/Q a brute force approach is used, where a range of cavities were simulated in order to build up a plot of the R/Q trends.

2.2.2 HOM Calculation - Sector Method

The mode identification method is a very slow and clumsy method of mode targeting. Much better would be to set the boundaries to the angle which contains only one maximum. Then the fundamental mode is the mode of interest, in this case the $TM_{10,1,0}$ mode, and an angle of 18° between the boundaries. MWS unfortunately does not allow symmetry planes to be placed at arbitrary angles, only $90^\circ/180^\circ$. A work around for this problem is to model only 18° of cavity with a symmetry plane, allowing only half of this volume to be meshed (see fig. 2.3). MWS's PBA³ helps to avoid errors introduced by the cavity side at an angle to the mesh cells. This is not appropriate without alteration for accurately calculating Q_0 ($\frac{\omega_0 W}{\Delta W}$ where W is the stored energy and ΔW is the power loss) values as the losses introduced by the extra conductors will affect the surface losses, although the surface materials could be altered to correct this. The R/Q found using this method is the R/Q as seen by the beams in a klystron.

Using this method is much more reliable and significantly faster as only one mode need be calculated.

³perfect boundary approximation

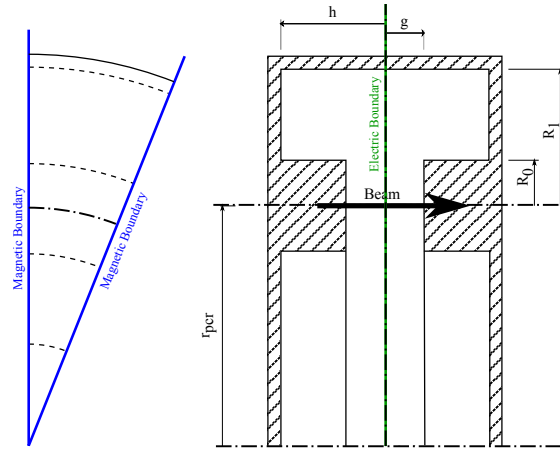


Figure 2.3: The boundary conditions used in MWS for the coaxial HOM cavity, whispering gallery cavities (both $TM_{10,1,0}$ modes).

Green: electric boundary, Blue: magnetic boundary

All boundaries in this case are electric boundaries.

The wedge shows the minimum amount of cavity that can be meshed while still representing the correct fields.

2.2.3 MBK R/Q Definition

There are several definitions of R/Q used by different people for different purposes. Two of particular interest are [64]:

$$\text{Tube engineer's definition } R/Q = \frac{|V_g|^2}{2\omega W} \quad (2.3)$$

$$\text{MWS } R/Q = \frac{|V_g|^2}{\omega W} \quad (2.4)$$

where $V_g = \int_{-\infty}^{\infty} E_g(0, z) dz$; $\omega = 2\pi f$; $W =$ stored energy.

Thales use the tube engineer's definition for their klys2D code, so use of MWS inbuilt definition would cause the simulations to overestimate the interactions. Fortunately MWS implements its R/Q calculations in a VBA⁴ script which can be altered so the definition of R/Q used can be changed to the tube engineers definition. For simulation of an MBK the differences do not stop there. Klys2D(see chap. 6) does not have a dedicated MBK mode and as such

⁴Visual Basic for Applications

all MBKs are simulated as a series of single beam klystrons. As there are several beams interacting with the same cavity field the definition of R/Q should be changed to ensure the klys2D takes this fact into account.

For the purposes of interaction simulation, each section of the cavity can be thought of as a separate cavity (although they are of course linked), MWS considers the full cavity and so calculates the R/Q incorrectly for these purposes.

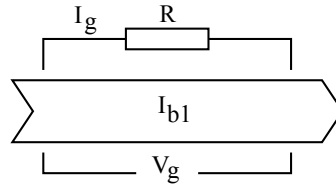


Figure 2.4: Simple model of interaction gap and beam with AC gap current I_g and AC voltage V_g ; cavity losses R_g and first harmonic of the beam current I_{b1}

Using a simple small signal model to represent the interaction gap, the AC gap voltage V_g can be defined in terms of I_g , gap coupling factor (M), Q_0 and R/Q (see fig. 2.4).

For a single beam in a full cavity (1 beam of N possible):

$$I_g = MI_{b1} \quad (2.5)$$

where I_{b1} is the first harmonic of the beam current.

using Ohm's law:

$$V_g = I_g R_g \quad (2.6)$$

and we can derive this in terms of R/Q as:

$$V_g = MI_{b1}(R/Q)_{x/N}Q_0 \quad (2.7)$$

where $(R/Q)_{x/N}$ is the R/Q seen by a beam in a cavity with x beams present and N possible beams. MWS calculates $(R/Q)_{1/N}$ whereas the quantity requi-

red by klys2D is $(R/Q)_{N/N}$ as it only simulates a single beam in a section of the cavity.

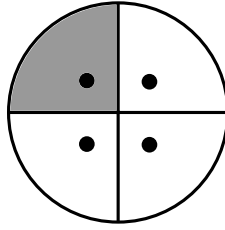


Figure 2.5: 4 beam klystron showing the stored energy interacted with by one beam (1/4 of the total energy). So the R/Q calculated for a single beam in a full cavity is in fact $1/N$ times the correct R/Q seen by the beams (due to the increased stored energy being associated with the beam).

For N beams of a possible N , the current and shunt impedance must be redefined. R_g is now derived from a stored energy of W/N as opposed to W as we only consider a section of the cavity. The R/Q calculated at one point $((R/Q)_{1/N})$ must be multiplied by the number of beams N to find the R/Q needed to simulate beams independently $((R/Q)_{N/N})$ (see eq. 2.3).

$$(R/Q)_{1/N} \cdot N = (R/Q)_{N/N} \quad (2.8)$$

In order to confirm this result the current and shunt impedance must be redefined. R_g and I_b are now derived from a stored energy of W/N as opposed to W , so:

$$R_g \rightarrow (R_g)_{1/N} / N \quad (2.9)$$

$$I_g \rightarrow I_g \cdot N \quad (2.10)$$

Substituting into eq. 2.6:

$$V_g = M(N \cdot I_g) \cdot R_g / N \quad (2.11)$$

So V_g is unchanged with respect to N as expected.

2.3 Re-entrant Cavity ($TM_{0,1,0}$)

2.3.1 Description

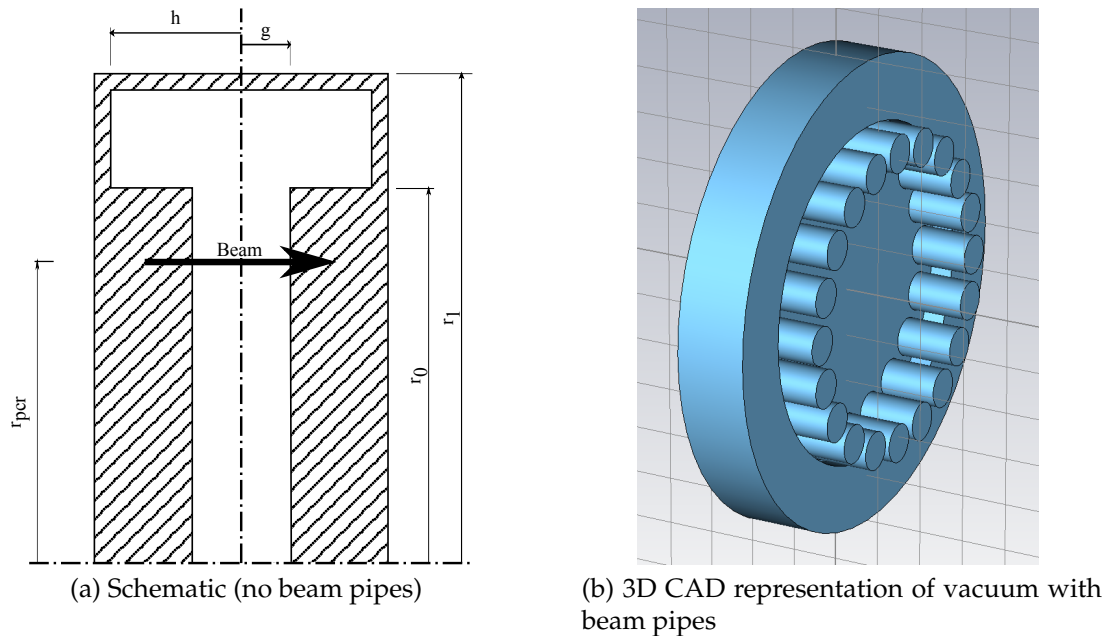


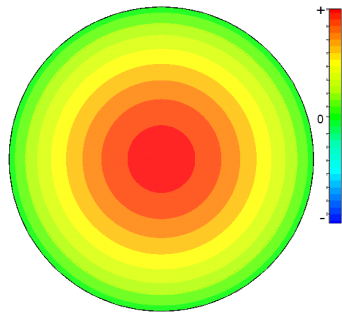
Figure 2.6: Re-entrant cavity

This is the standard choice of cavity for single beam klystrons and was used in THALES' TH1801[55].

The electric field concentration at the centre of the cavity forms its greatest limitation when considered for use in an MBK. The TH1801 attempts to solve this problem by using a central beam with 6 beams surrounding it. This leads to a fairly compact configuration where the field at the outer beams is still fairly high. However as the central beam differs from the outer beams optimisation of the structure for all beams is non trivial.[29]

The cavity in this case must be relatively large to accommodate the 20 beams mandated by the design (see sec. 1.4.2). This leaves the beams far from the field maximum in the centre limiting the R/Q and so the peak voltage seen by the beams.

2.3.2 Initial Parameter Results



Parameter	Value(mm)
g	5.3
r_{pcr}	62.8
r_0	$r_{pcr} + 9.2$
r_1	73-90
h	11-70

Table 2.2: Re-entrant cavity: Dimensions (see table. 2.1)

Figure 2.7: Re-entrant cavity: E_z plot in radial plane of $TM_{0,1,0}$ Mode

Using the brute force approach to cover the parameter space fully (see fig. 2.8), the 1.3 GHz line follows the R/Q contours well when h is less than 25 mm. This gives good flexibility when selecting an appropriate cavity geometry so other factors (such as multipactor) can be taken into account when choosing the optimum geometry.

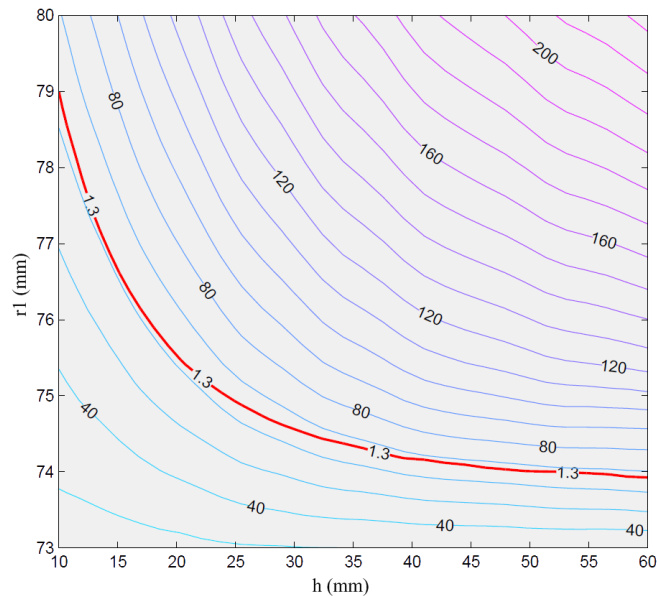


Figure 2.8: Re-entrant Cavity: R/Q contours plotted against h & r_1 (plus 1.3 GHz frequency contour) showing maximum R/Q of $\sim 70 \Omega$

Table 2.3: Re-entrant cavity: Optimal parameters and simulation results

Parameter	Value	Units
r_1	77.44	mm
h	15	mm
R/Q	68.3	Ω
Q	4242.67	
f	1.301	GHz

Results of the optimal simulation (see table. 2.3) are what would be expected from fig. 2.8 and confirm that the highest R/Q that can be expected within the confines of our parameters is around $\approx 68.3 \Omega$.

2.4 Recessed Re-entrant Cavity ($TM_{0,1,0}$)

2.4.1 Description

The main weakness of the re-entrant cavity in this configuration concerns the beam interaction with the stored energy. The beams only interact with the edge of the E-field leaving the maximum in the centre and its associated stored energy untouched. The natural way to solve this problem is to recess the centre of the cavity to concentrate the E-field around the beams using a ridge. This should aid interaction with the field and so increase the R/Q. For the ideal optimisation the cavity should have no field in the centre.

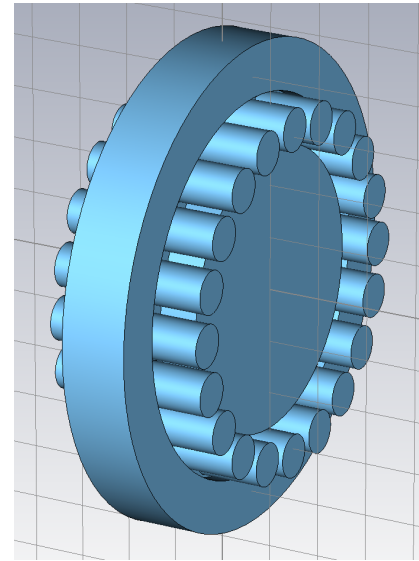
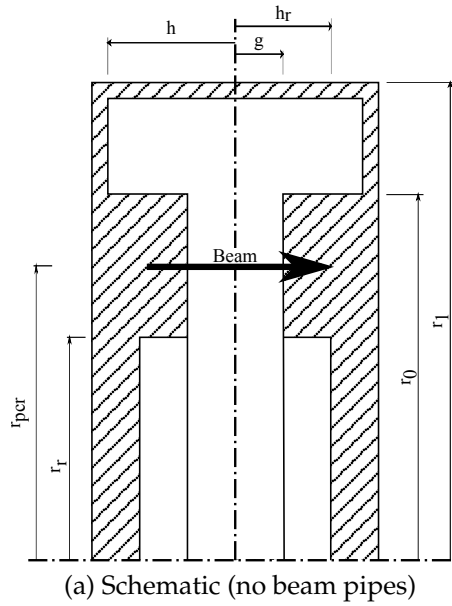
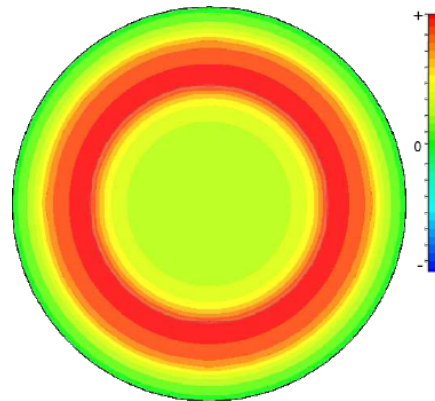


Figure 2.9: Recessed re-entrant cavity

Table 2.4: Recessed re-entrant cavity: Dimensions (see table. 2.1)

Parameter	Value (mm)
g	5.3
r_{pcr}	62.8
r_0	9.2
h_r	5.3-150
r_1	70-115
h	5.3-40

Figure 2.10: Recessed re-entrant cavity: E_z plot in radial plane of $TM_{0,1,0}$ Mode

2.4.2 Initial Parameter Results

To characterise the parameter space for a cavity with three free dimensions a 3D graph must be used. Fig. 2.11 achieves this by interpolating between the points in the 3D parameter space to find a surface with a frequency of 1.3 GHz.

On to this is plotted a colour map representing the interpolated R/Q expected for every point on the surface.

The 3D surface found can help to find the optimal R/Q, but the point found is imprecise due to the nature of a colour map. It can with some confidence be used to optimise one of the parameters loosely, in this case h_r (the depth of the recess) lends itself to this approach (see fig. 2.12). In order to make use of existing simulation data, h_r was chosen to lie on an existing plane of simulations in the 3D parameter space. This plane can now be rendered as a 2D contour plot as in the case of the re-entrant cavity.

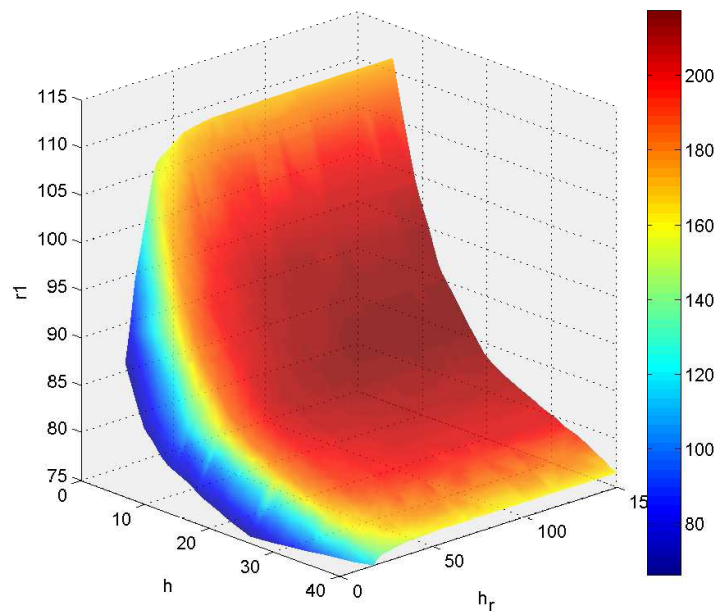


Figure 2.11: Recessed re-entrant cavity: Iso-frequency surface where $f=1.3$ GHz. R/Q is represented by the colour map, showing a maximum of ~ 220 Ω

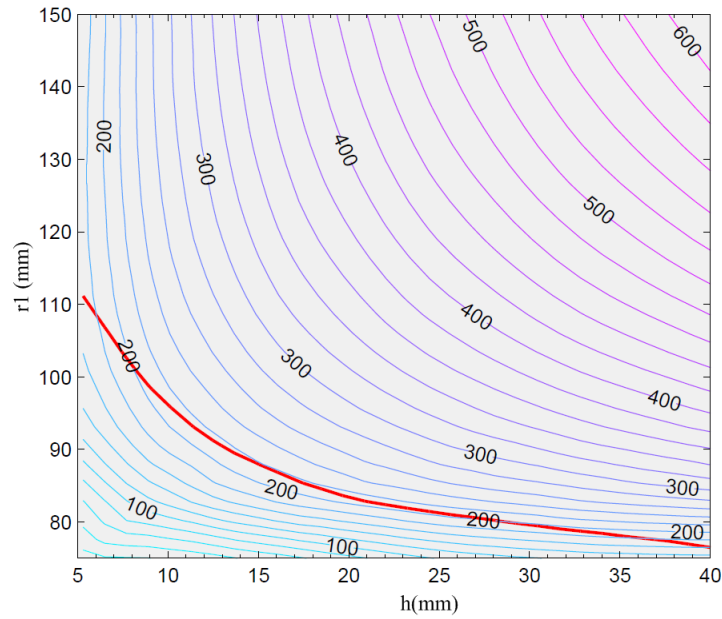


Figure 2.12: Recessed re-entrant cavity: R/Q contours plotted against h & r_1 (plus 1.3 GHz frequency contour) where $h_r = 101.77$ mm

Table 2.5: Recessed re-entrant cavity: Optimal parameters and simulation results

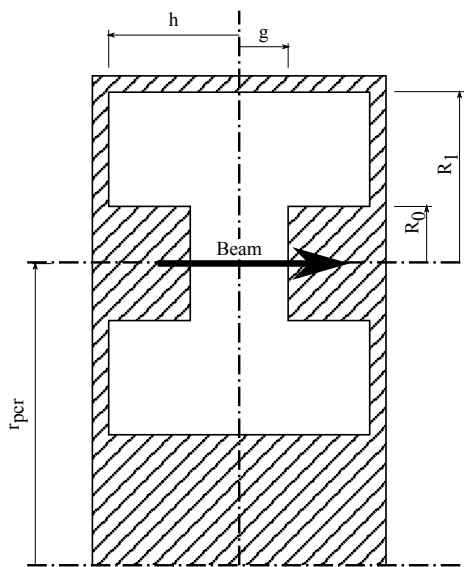
Parameter	Value	Units
r_1	86.9	mm
h	16	mm
h_r	101.77	mm
r_{pcr}	62.8	mm
R/Q	216.1	Ω
Q	6200.38	
f	1.292	GHz

From fig. 2.11 it can be seen that once h_r is in the region of 70 mm-100 mm any further increase leaves the R/Q roughly stable. In fig. 2.12 h_r is fixed to 101.77 mm in order to more accurately identify the highest R/Q. Again a range of values would be appropriate, allowing the cavity size to be determined by other factors. The R/Q of 216 Ω is much higher than for the re-entrant cavity (see sec. 2.3) as would be expected.

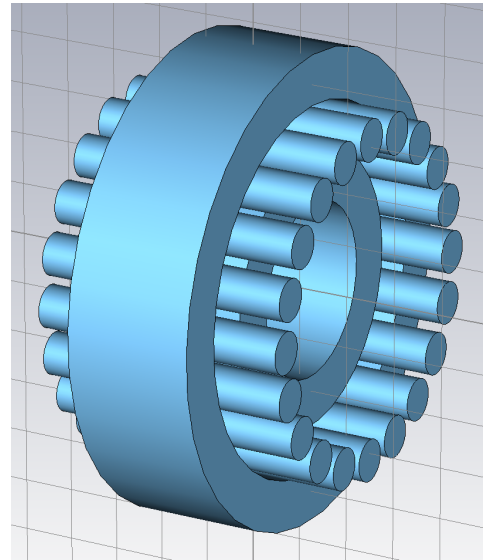
Although a large h_r yields the highest R/Q it restricts the choice of interaction structure as the distance between the gaps will be forced to be large. A smaller choice of h_r for instance 30 mm would still yield an R/Q of around 180 Ω and would allow much greatest flexibility in cavity placement.

2.5 Fundamental Mode Coaxial Cavity ($TM_{0,1,0}$)

2.5.1 Description



(a) Schematic (no beam pipes), note R_0 and R_1 are radial distances, not radii



(b) 3D CAD representation of vacuum with beam pipes

Figure 2.13: Coaxial cavity

In the recessed re-entrant cavity (see sec. 2.4) the centre of the cavity is recessed to concentrate the electric field on the beam ridge.

In this vein the centre of the cavity could be removed all together leaving a toroidal geometry. This avoids the large bulge in the centre of the cavity associated with the recessed re-entrant cavity and eases use in a practical tube.

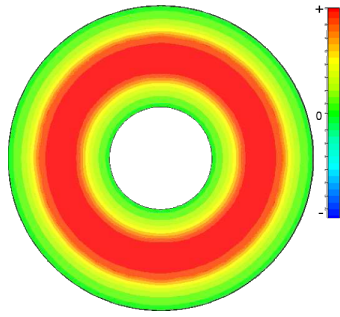


Figure 2.14: Fundamental mode coaxial: E_z plot in radial plane of $TM_{0,1,0}$ Mode

Table 2.6: Fundamental mode coaxial: Cavity dimensions (see table. 2.1)

Parameter	Value (mm)
g	5.3
R_0	9.2
r_{pcr}	62.8-100
R_1	11-60
h	11-50

2.5.2 Initial Parameter Results

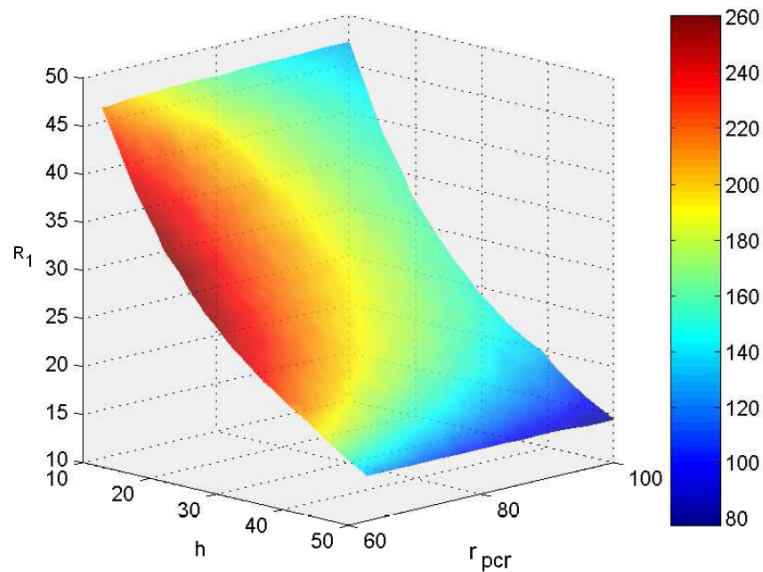


Figure 2.15: Fundamental mode coaxial: Iso-frequency surface at 1.3 GHz. R/Q is represented by the colour map

In fig. 2.15 the R/Q is clearly highest when the r_{pcr} is 62.8 mm, its minimum as limited by the size of the cathodes. The selection of the optimal geometry is therefore better represented by fig. 2.16 where a slice is taken at $r_{pcr} = 68.2$ mm. This clearly shows the position of the optimal R/Q. There is limited flexibility in

this design, although its high potential R/Q (260 Ω) means that a non optimal geometry could be used while still yielding an appropriately high R/Q.

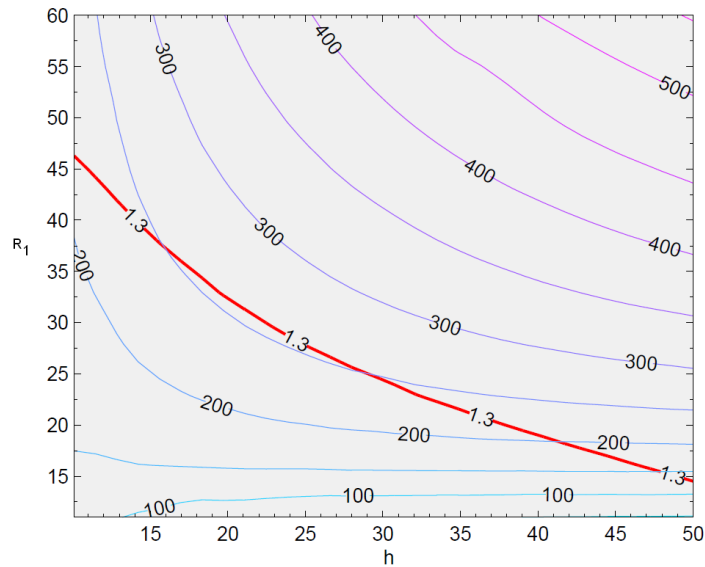


Figure 2.16: Fundamental mode coaxial: R/Q contours plotted against h & r_1 (plus 1.3 GHz frequency contour) for coaxial cavity ($r_{pcr} = 62.8$)

Table 2.7: Fundamental mode coaxial: Optimal parameters and simulation results

Parameter	Value	Units
R_1	31.6	mm
h	21	mm
r_{pcr}	62.8	mm
R/Q	256.63	Ω
Q	7939.39	
f	1.293	GHz

This geometry does not increase the potential R/Q by a large amount over the recessed re-entrant cavity but it does have the advantage of being far more compact, because of this it lends itself to use in a practical tube.

2.6 HOM Coaxial Cavity ($TM_{10,1,0}$)

2.6.1 Description

The fundamental mode of the coaxial cavity has the electric field tightly concentrated around the beams. This could perhaps be further optimised by using the $TM_{10,1,0}$ mode. With this mode each beam interacts (in anti phase compared to its neighbours) with a maximum. An increased R/Q may be found due to the advantageous distribution of stored energy, but in order to tune the cavity the r_{pcr} will have to be greatly increased having the opposite effect. The $TM_{10,1,0}$ (see fig. 2.17) is used because of the design decision of 20 beams (see sec. 2.9.6). The cavity's construction remains identical to the standard coaxial cavity to ease comparison (see sec. 2.5).

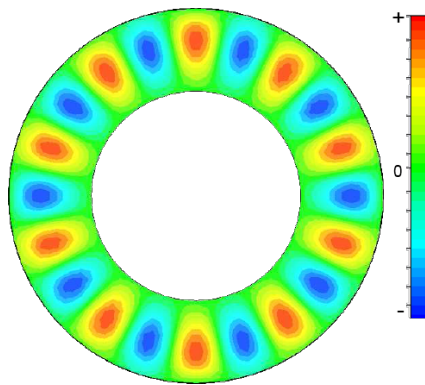


Table 2.8: HOM coaxial cavity: Cavity dimensions (see table. 2.1)

Parameter	Value (mm)
g	5.3
r_{pcr}	350-500
R_0	9.2
h	5.3-100
R_1	20-150

Figure 2.17: HOM coaxial cavity: E_z plot in radial plane of $TM_{10,1,0}$ mode

2.6.2 Initial Parameter Results

All of the simulations on this geometry were performed using the segment method due to its markedly lower calculation time.

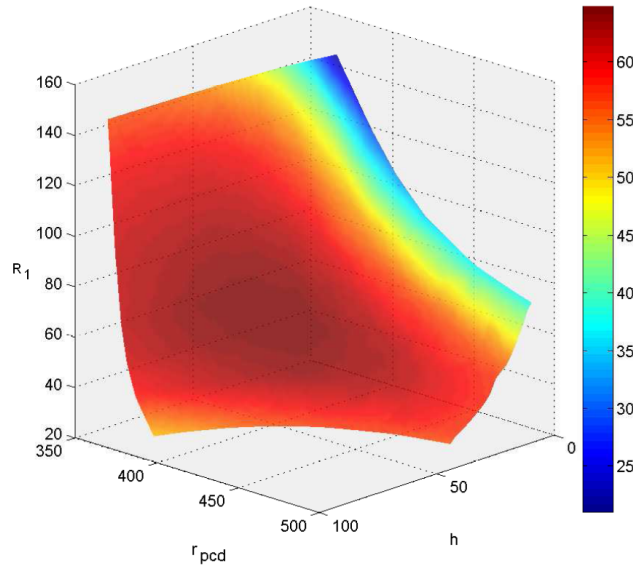


Figure 2.18: HOM coaxial cavity: Iso-frequency surface at 1.3 GHz. R/Q is represented by the colour map

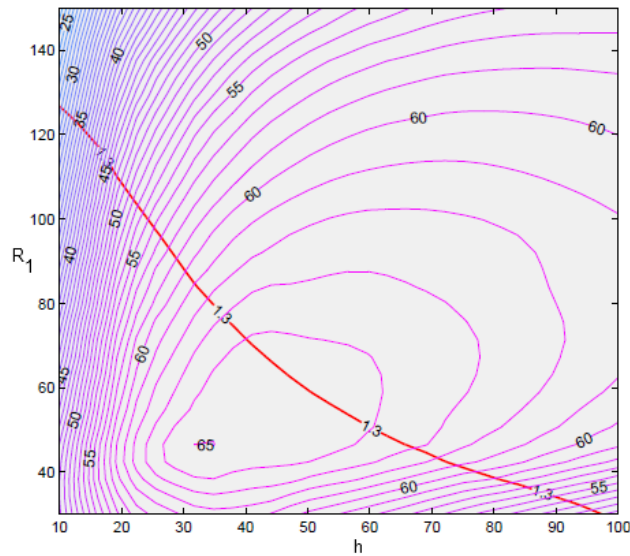


Figure 2.19: HOM coaxial cavity: $r_{pcr} = 400$ mm

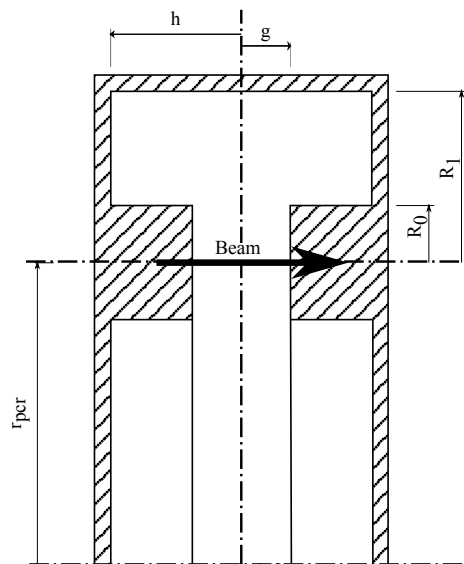
Using fig. 2.18 which is centred around a r_{pcr} of 400 mm (a figure obtained from fig. 2.19) an optimum geometry can be easily identified (see table. 2.9). The optimum R/Q is of the order of 65Ω which is approximately the same as for the re-entrant cavity (see sec. 2.3).

Table 2.9: HOM coaxial cavity: Optimal parameters and simulation results

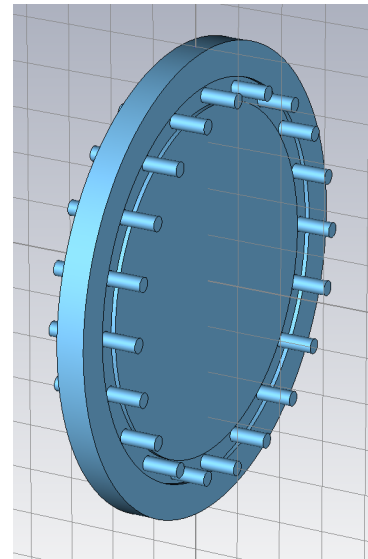
Parameter	Value	Units
R_1	61.59	mm
h	48	mm
r_{pcr}	400	mm
R/Q	64.5	Ω
Q	5526.98	
f	1.301	GHz

2.7 Whispering Gallery Cavity ($TM_{10,1,0}$)

2.7.1 Description



(a) Schematic (no beam pipes), note R_0 and R_1 are radial distances, not radii



(b) 3D CAD representation of vacuum with beam pipes

Figure 2.20: Whispering gallery cavity

The whispering gallery cavity is a further refinement of the recessed re-entrant cavity (see sec. 2.4) or a simplification of the HOM coaxial cavity (see sec. 2.6) dependant on your point of view. The whispering gallery cavity uses the same mode as the HOM coaxial cavity (see sec. 2.6) to attempt to focus the electric field on the beam locations. However this cavity could experience the same low R/Q as the HOM coaxial cavity.

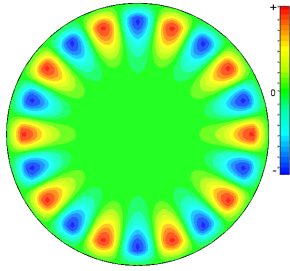


Figure 2.21: Whispering gallery cavity: E_z plot in radial plane of $TM_{10,1,0}$ Mode shape in the whispering mode cavity

Table 2.10: Whispering gallery cavity: Cavity dimensions (see table. 2.1)

Parameter	Value (mm)
g	5.3
R_0	9.2
h	10-40
r_{pcr}	100-500
R_1	100-550

Again the $TM_{10,1,0}$ is used because of the design decision of 20 beams 1.4.2.

2.7.2 Initial Parameter Results

All of the simulations on this geometry were performed using the segment method due to its markedly lower calculation time.

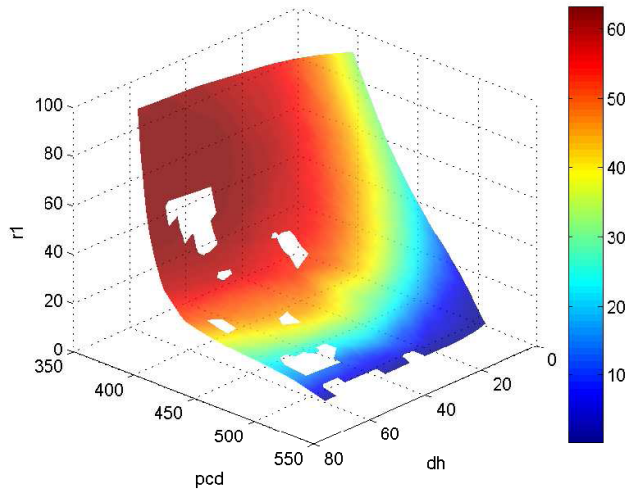


Table 2.11: Whispering gallery cavity: Optimal parameters and simulation results

Parameter	Value	Units
R_1	74.07	mm
h	55.76	mm
r_{pcr}	388.3	mm
R/Q	62.7	Ω
Q	5480.6	
f	1.290	GHz

Figure 2.22: Whispering gallery cavity: Iso-frequency surface at 1.3 GHz. R/Q is represented by the colour map for whispering gallery cavity with beam ridge simulated in a segment. The gaps in the plot are an artifact of the interpolation methods used in MatLab.

The R/Q achieved here is comparable to the re-entrant cavity (see sec. 2.3) and the HOM coaxial cavity (see sec. 2.6).

2.8 Preliminary Conclusions

As expected the re-entrant cavity did not perform well due to the large amount of stored energy which the beams do not interact with. However the refinement of the recessed re-entrant cavity successfully solves this problem and displays promising performance. The length indicated in Table 2.12 is in this case merely an optimum, past which no improvement is made. It could be appreciably shortened with only a small loss in performance.

The fundamental mode coaxial cavity gives the advantages of the recessed re-entrant cavity but with a reduced length and a higher mechanical strength. The higher order mode cavities prove to be inferior as far as R/Q is concerned.

They do however allow the radius of the tube to be increased appreciably, which could be of great importance with respect to cooling and cathode loading when the CLIC specification is used. However this increase in r_{pcr} is at the expense of R/Q meaning that extended interaction gaps or other such structures may be required to make the tube operate satisfactorily.

Table 2.12: Initial investigation conclusions

	1.3 GHz Resonance				
	Re-entrant	Recessed Re-entrant	Coaxial $TM_{0,1,0}$	Coaxial $TM_{10,1,0}$	Whispering Gallery
R/Q (Ω)	68.3	216.1	256.3	64.8	62.7
Q	~4000	~6000	~8000	~5500	~5500
Outside radius r_1 (mm)	77.44	86.9	94.4	461.59	462.07
R_1 (mm)			31.6	61.59	74.07
r_{pcr} (mm)	62.8	62.8	62.8	400	388
r_t (mm)	8.21	8.21	8.21	8.21	8.21
r_0 (mm)	$r_{pcr} + 9.2$	$r_{pcr} + 9.2$			
R_0 (mm)			9.2	9.2	9.2
Height (h) (mm)	15	16 ($h_r \sim 100$)	21	48	55.76
Damping	possible	easy	possible	possible	easy

Out of the cavities investigated, when R/Q and convenience are considered, the $TM_{0,1,0}$ coaxial cavity is the best choice although the recessed re-entrant cavity has merit. The $TM_{10,1,0}$ based cavities simply do not provide a large enough R/Q to be used in high efficiency interaction structures.

2.9 Revised Parameters

The initial cavity study section uses a set of preliminary parameters to characterise the cavities. Results from the interaction simulations and discussion with Thales led to the conclusions that some of the klystron parameters required revision (see chap. 6).⁵ High cathode loading for instance, leads to low convergence, which would make a practical focusing system and gun hard to

⁵Private discussion with Thales Electron Devices Technical Staff

design, especially around the gun region. This difficulty is also due to the relatively large beam. This would be exacerbated by the small cathode spacing placing the beams too close together, which additionally could have lead to thermal issues in the beam pipe noses. The interaction simulations also showed that the gap was in danger of voltage breakdown when run at the target power (see table. 2.13).

Table 2.13: Preliminary klystron parameter problems

Short output gap(see chap. 6)	10.6 mm
High cathode loading	5.5 A/cm ²
Insufficient cathode separation	3 mm
Excessively low area convergence	2

The preliminary parameters are revised in this section (see table. 2.14).

Table 2.14: Summary of preliminary fixed klystron/cavity parameters

Property	Preliminary
gap length (g)	10.6 mm
beam radius b (r_b)	5.74 mm
tunnel radius a (r_t)	8.21 mm
cathode loading (j_c)	5.5 A/cm ²
cathode separation (d_{cs})	3 mm
beam ridge width (r_0)	9.2 mm
pitch circle radius (r_{pcr})	62.8 mm
b/a	0.7

2.9.1 Cavity Gap Length

The standard interaction gap of the cavities has been lengthened from 10.6 mm to 16 mm (with $\beta_e g = 0.75$ which is below the limit 1) to help avoid voltage breakdown in the output gap. A promising interaction structure using the recessed re-entrant/fundamental mode coaxial cavity was created using Klys2D where the output gap was on the limit of safe voltage gradients (at ~ 11 kV/mm). The longer interaction gap should help to solve this problem.

2.9.2 Radius of Cathode Circle (r_{pcr})

The r_{pcr} is the radius of a circle on which the centres of the beams lie. This is defined by the beams size and the distance between cathodes. It defines (along with the cathode size) the minimum radial size of the cavities as the beams must fit inside them. A balance has to be struck between the R/Q of the cavities (which increases as their radius decreases) and the ability to design and build adequate guns. Initially a very low cathode spacing was chosen to maximise the R/Q which is very small compared to the THALES 1801 7 beam MBK.[29] This spacing is used in the initial cavity study as evidenced by the small cavities.

In the light of this comparison the spacing was increased to be in line with the TH1801 in order to ensure the practicality of the tube.

If the cathode separation is increased to 10 mm (using the same cathode loading 5.5 A/cm^2) this increases the r_{pcr} from 62.8 mm to 107.39 mm. This is much larger than before and may cause a lower maximum R/Q in the fundamental mode cavities.

In the initial cavity study a cathode loading of 5.5 A/cm^2 was used, which is quite high even with the possibility of new Thales cathode technology.⁶ High cathode loading leads to reduced tube lifetime, due to premature failure of the electron guns. This is because of barium evaporation from the cathode surface. Lower loading can lead to greatly increased lifetime. In a large and expensive tube with multiple cathodes this is clearly a great advantage.

The r_{pcr} is found from cathode loading by first calculating the cathode radius using eq. 1.4. So the radius of the minimum circle which can contain n_b beams can be simply found:

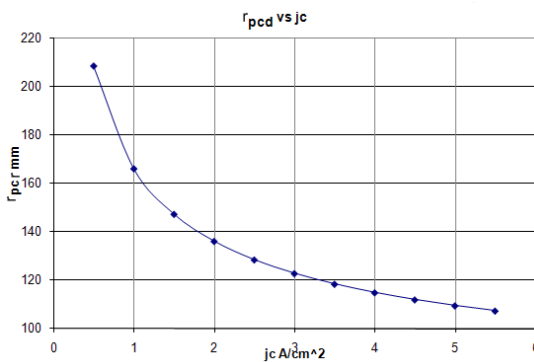
$$r_{pcd} = \frac{(r_c + d_{cs}) \cdot N_b}{\pi} \quad (2.12)$$

where d_{cs} = cathode spacing; r_{pcr} = pitch circle radius; N_b = number of beams

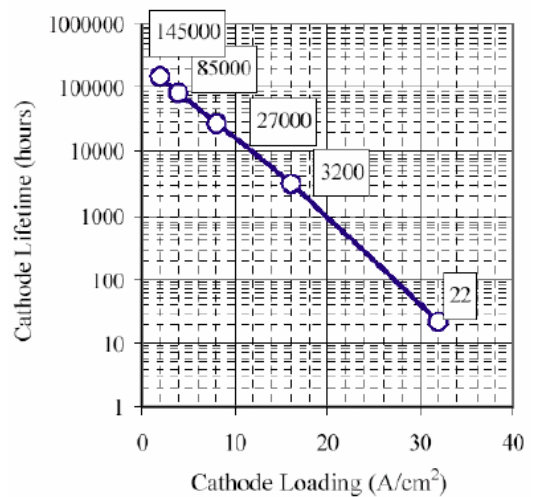
⁶Private discussion with Thales Electron Devices Technical Staff

A cathode loading of 3.5 A/cm^2 giving a lifetime of over 10 years (~ 85000 hours) is a good compromise between lifetime and r_{pcr} (see fig. 2.23). A shorter lifetime would lead to frequent replacement as the experiment is due to last 10 years and a longer life time would result too low a current density and too large a cathode. A lower cathode loading also has the effect of increasing the low convergence to a more suitable value by increasing the cathode surface area to retain a constant current.

From fig. 2.23 it can be seen that the new r_{pcr} is 118 mm from an initial value of 62.8 mm.



(a) r_{pcd} against J_c (cathode loading) including 10 mm cathode spacing



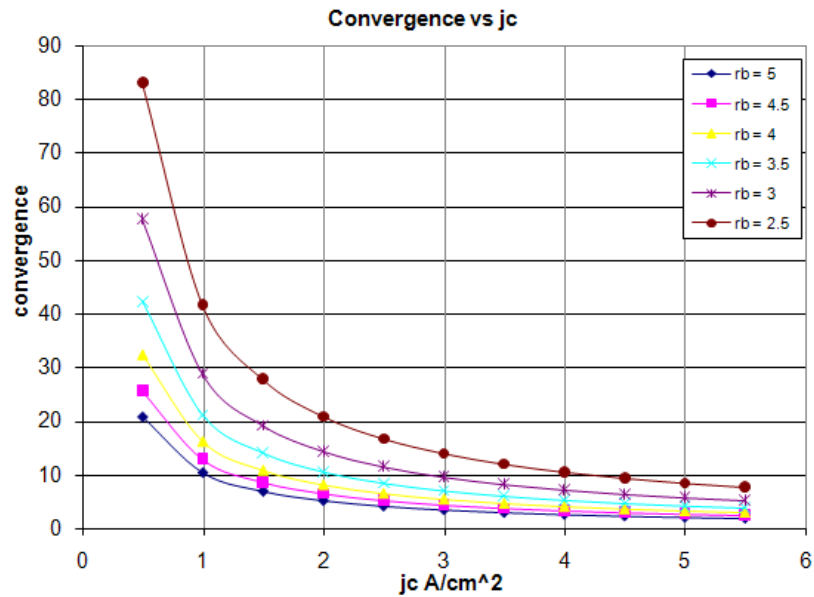
(b) Cathode lifetime w.r.t. cathode loading, picture taken from [40]

Figure 2.23: 3.5 A/cm^2 is a good compromise between lifetime and overall tube radius.

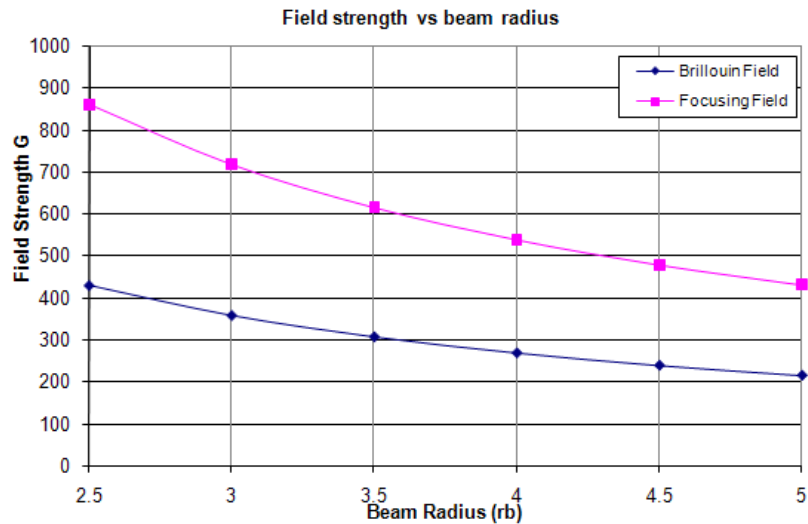
2.9.3 Area Convergence

The initial investigation used a very low convergence of around 2. This was caused by the small cathode radius allowed by the high cathode loading. As the cathode loading has been lowered the convergence will naturally increase but we can also change the beam radius. A higher convergence is required for A good choice of beam radius might be 3.5 mm when combined with our chosen cathode loading (see fig. 2.24).

The focusing field is calculated using the definition of the Brillouin field and focusing field in eq. 1.8. This value also strikes a balance between convergence and focusing field, giving an area convergence (see eq. 1.2) of 6 as opposed to an initial value of 2 (see eq. 1.2) with the new value being more suitable for an effective electron gun design ⁷.



(a) Convergence against J_c (cathode loading)



(b) Field strength against beam radius with a Brillouin field ratio (B_n) of 2

Figure 2.24: The definition of cathode loading allows the selection of the beam radius with consideration of the area convergence of the gun and the focusing field strength required.

⁷Private discussion with Thales Electron Devices Technical Staff

2.9.4 Tunnel to Beam Ratio (b/a)

The initial investigation used a tunnel to beam ratio (b/a) of 0.7, for the further investigation b/a has been changed to 0.6 in line with tubes designed at Thales.⁸

So with the new beam radius the beam and tunnel parameters are:

b/a	beam radius (b)	tunnel radius (a)
0.6	3.5 mm	5.8 mm

2.9.5 Beam Ridge Width (R_0)

The cavity models studied in chapter 2 all have the beam pipe noses combined as a ridge. The width of this ridge is $2 \cdot r_0$ with its centre line at r_{pcr} . This is a critical dimension as ideally it would be minimised to increase the R/Q, however it must also be large enough to allow cooling pipes to be embedded to avoid high temperatures in the beam pipe noses. In this initial investigation r_0 was kept to a minimum to fully optimise the R/Q, however this minimises the amount of copper in the beam pipe noses and could cause problems with over heating. The minimum thickness of the copper is radius of the tunnel (a) minus beam ridge width r_0 . The cavities are simulated with a range of r_0 from 7-21 mm or a wall thickness of 1.2 mm to 15 mm.

2.9.6 Final Parameters

The r_{pcr} has been increased sufficiently to allow a more conservative cathode loading and separation and thus increased the convergence. The convergence has also been increased by the reduction of the beam (and tunnel radius). The gaps have also been extended to ease high output gap voltage gradient. The parameters are summarised in table. 2.15. The new arrangement of beams and cathodes shows the sparser layout of beams and the circle on which they lie (see fig. 2.25).

⁸Private discussion with Thales Electron Devices Technical Staff

Table 2.15: Summary of changes in the revised parameters

Property	Preliminary	Revised
gap length (g)	10.6 mm	16 mm
beam radius (b)	5.74 mm	3.5 mm
tunnel radius (a)	8.21 mm	5.835 mm
cathode loading (j_c)	5.5 A/cm ²	3.5 A/cm ²
cathode separation (d_{cs})	3 mm	10 mm
beam ridge width (r_0)	8.21 mm	7 mm to 21 mm
b/a	0.7	0.6
pitch circle radius (r_{pcr})	62.8 mm	118.5 mm
required focusing field (factor of 2 above the Brillouin field)	370 G	600 G
convergence	2	6

The larger r_{pcr} (and so larger cavities) and greater focusing field may rule out using a solenoid for focusing. One of the advantages of the fundamental mode cavities over the whispering gallery cavity was their relatively small size, comparable to the TH1801.[29] This suggested the possibility of using existing focusing technology.

These revised parameters are used to re-evaluate the promising cavities in the second section of the cavity study and for the remainder of this thesis.

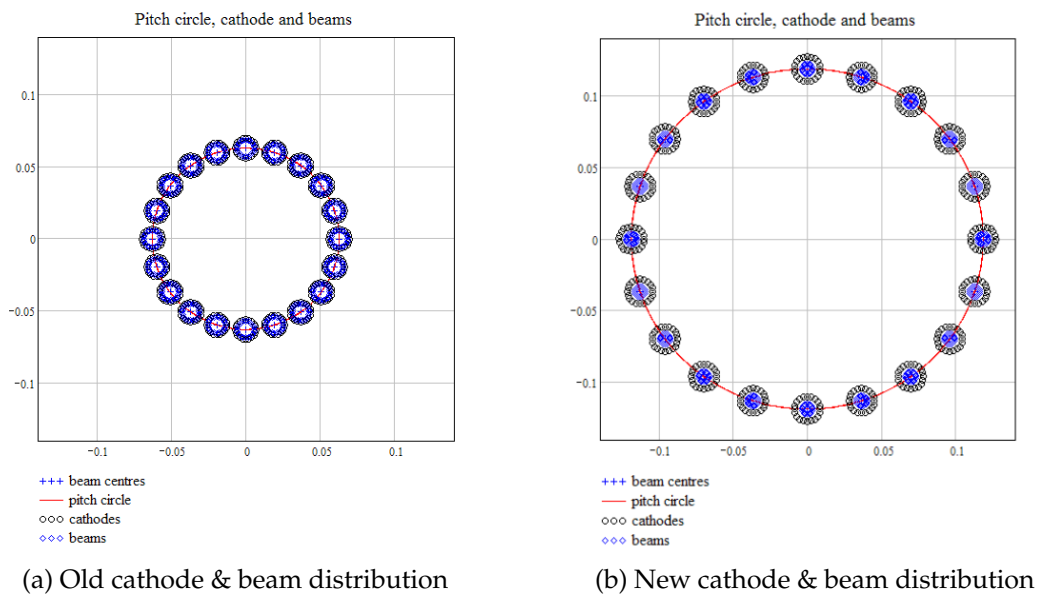


Figure 2.25: A comparison between the new and old beam/cathode arrangements

2.10 Recessed Re-entrant Cavity ($TM_{0,1,0}$)

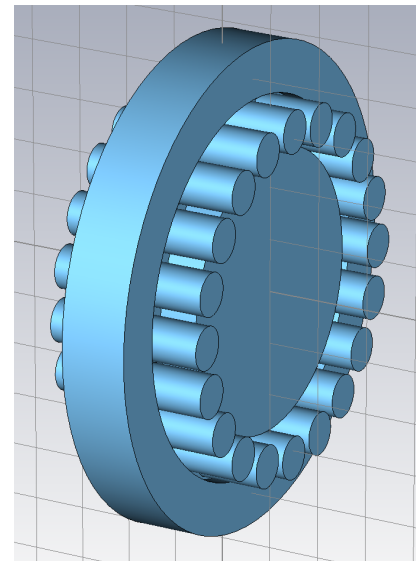
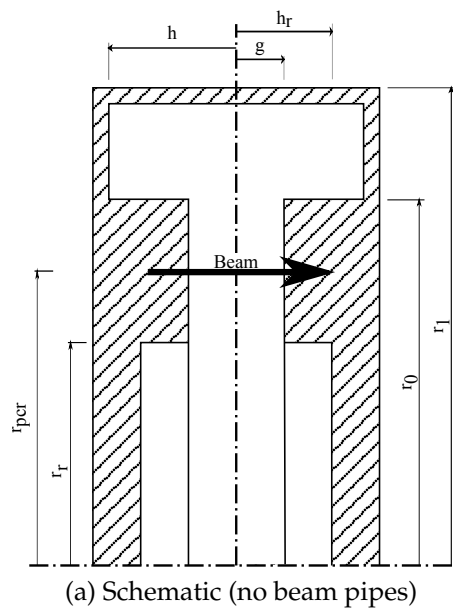


Figure 2.26: Recessed re-entrant cavity (repeated from Sec. 2.4)

2.10.1 Revised Parameter Results

The results from the initial investigation suggest that this cavity may be able to support a high R/Q fundamental mode. Unfortunately the recessed re-entrant cavity cannot be adjusted within the bounds of the revised klystron parameters to support a fundamental mode with a frequency of 1.3 GHz.

Table 2.16: Initial investigation recessed re-entrant cavity results (see table. 2.3)

(a) Initial investigation conclusions

	1.3 GHz Resonance
	Recessed Re-entrant
R/Q (Ω)	216.1
Q	~6000
Outside radius r_1 (mm)	86.9
r_{pcr} (mm)	62.8
r_t (mm)	8.21
r_0 (mm)	$r_{pcr} + 9.2$
Height (h) (mm)	16 ($h_r \sim 100$)
Damping	easy

Starting from a minimum of h which is enforced by the value of g (the cavity height cannot be smaller than the gap length) and increasing to an arbitrary maximum, it is clear that the trend in frequency is downwards. As the initial frequency is the highest value (1.018 GHz), this means that h must be minimised (preferably to g) which will have a detrimental effect on the R/Q as the small change in cavity height at the gap will not act to concentrate the field so effectively (see fig. 2.27).

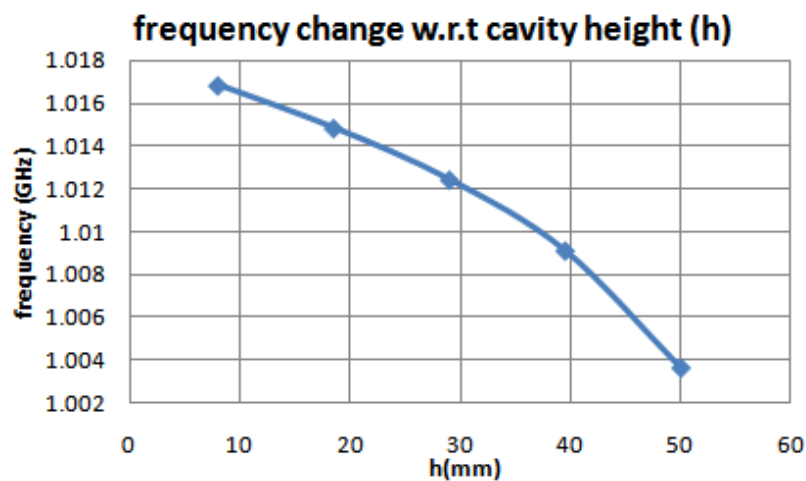


Figure 2.27: Frequency change of fundamental mode in GHz w.r.t. the height of the cavity. Increasing beyond its lower limit does not tend towards 1.3 GHz.

From a minimum of r_1 which is enforced by a r_{pcr} & r_0 (the cavity cannot be smaller than the beam ridge) and increasing again to an arbitrary maximum, the frequency trend is again downwards. Again the preferable value of r_1 is the minimum (see fig. 2.28).

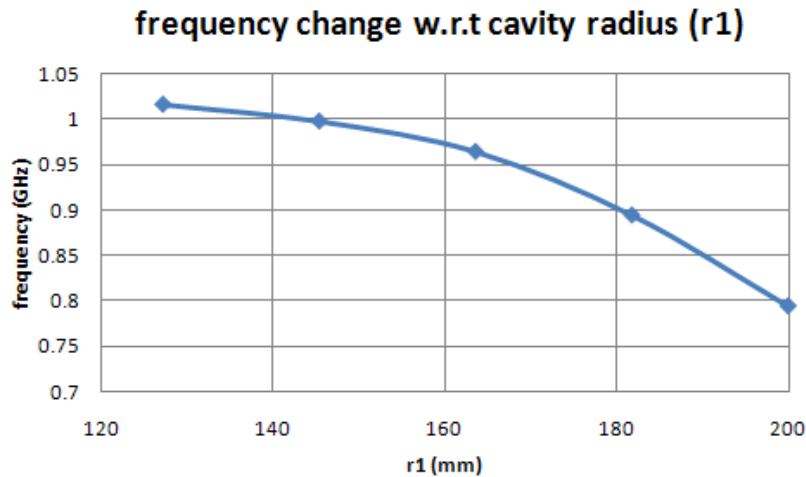


Figure 2.28: Frequency change of fundamental mode in GHz w.r.t. the outside radius of the cavity. Increasing beyond its lower limit does not tend towards 1.3 GHz.

Using the minima of h and r_1 and plotting from a minimum of h_r which is enforced by a r_{pcr} & r_0 (the recess in the centre of the cavity has no beneficial effect if it is smaller than the gap length) and increasing again to an arbitrary maximum, the frequency trend this time is upwards. Unfortunately the frequency maximum is still only 1.02 GHz and has begun to plateau almost 25% from the target frequency (see fig. 2.29).

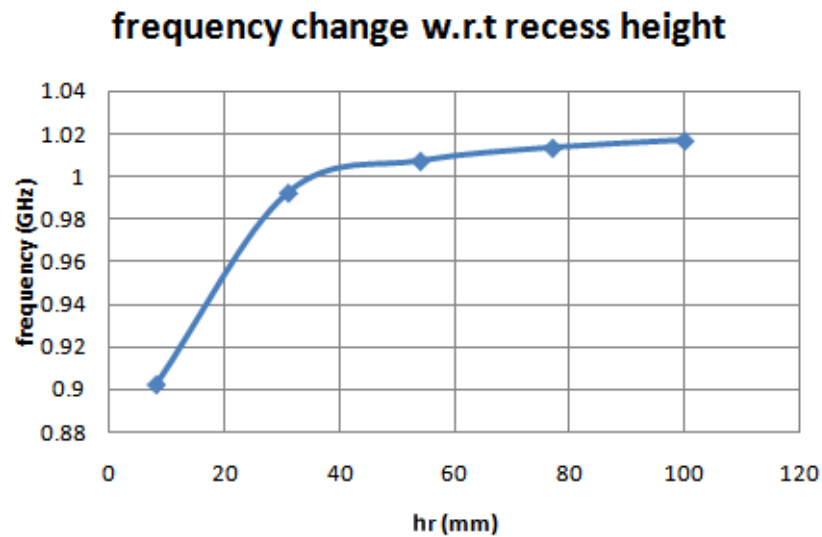
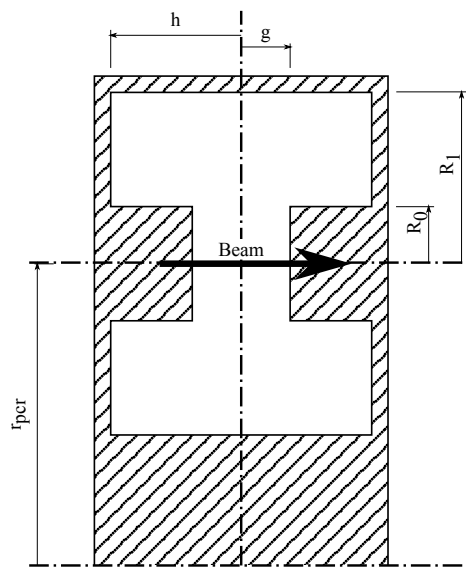


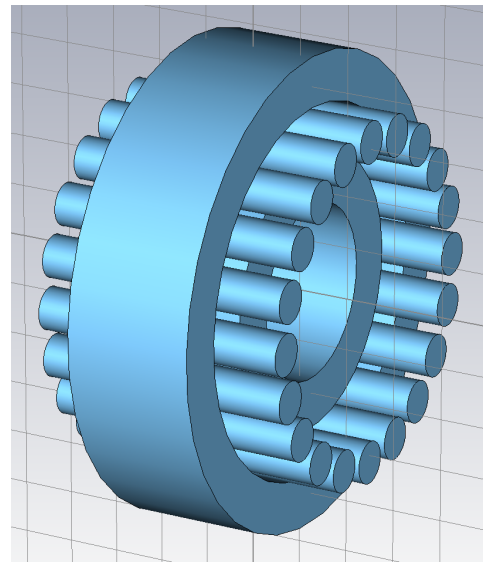
Figure 2.29: Frequency change of fundamental mode in GHz w.r.t. the height of the recess. Increasing beyond its lower limit the frequency tends towards 1.3 GHz but not before the cavity becomes too large.

Although the recessed re-entrant cavity looked promising from the initial investigations within the constraints imposed by the revised klystron parameters, it is no longer possible to find a cavity which supports a fundamental resonant mode of 1.3 GHz. This rules out this cavity for an MBK of this specification.

2.11 Fundamental Mode Coaxial Cavity ($TM_{0,1,0}$)



(a) Schematic (no beam pipes), note R_0 and R_1 are radial distances, not radii



(b) 3D CAD representation of vacuum with beam pipes

Figure 2.30: Coaxial Cavity (repeated from Sec. 2.5)

2.11.1 Revised Dimensions

The fundamental mode coaxial cavity has a similar performance to the recessed re-entrant cavity while being more longitudinally compact in comparison. An R/Q of 250Ω made it the favourite in that respect.

An altered optimisation method was employed for this repeated optimisation. As the frequency of a simulation converges with lower mesh density than the R/Q , a coarse low mesh parameter sweep was used to find the 1.3 GHz contour. With a polynomial fitted to this, a range of cavities with 1.3 GHz resonance can be simulated with higher mesh. This method is explained more fully in appendix. A.2.

Table 2.17: Initial investigation fundamental mode coaxial cavity results

	1.3 GHz Resonance
	Coaxial $TM_{0,1,0}$
R/Q (Ω)	256.3
Q	~8000
Outside radius r_1 (mm)	94.4
R_1 (mm)	31.6
r_{pcr} (mm)	62.8
r_t (mm)	8.21
R_0 (mm)	9.2
Height (h) (mm)	21
Damping	possible

The coaxial cavity in this case only has 2 free dimensions:

	Range
h	15-50 mm
R_1	15-50 mm

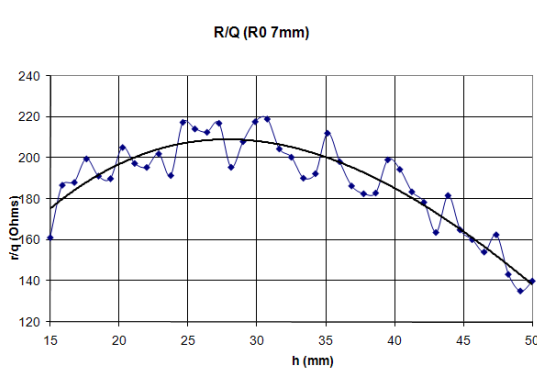
This range was chosen from a coarse parameter sweep which demonstrated that it covered the R/Q maxima adequately. For the base⁹ data set 15 steps were used in each direction giving a full set of 225 simulations to characterise the frequency space. For the line¹⁰ data set 40 simulations were used to show the trend of the R/Q. The polynomial function representing the 1.3 GHz cavities required 7 coefficients (optimised by overlaying the polynomial over the 1.3 GHz contour) (see sec. A.2).

Using the method described above (see sec. A.2) the range of cavities was calculated. This was repeated for a number of values of R_0 . Interestingly the maximum R/Q moves closer to the minimum h as R_0 increases. This is perhaps because as the capacitance C increases (due to larger R_0 and thus a larger beam ridge) the inductance L must be lowered accordingly by lowering the height of

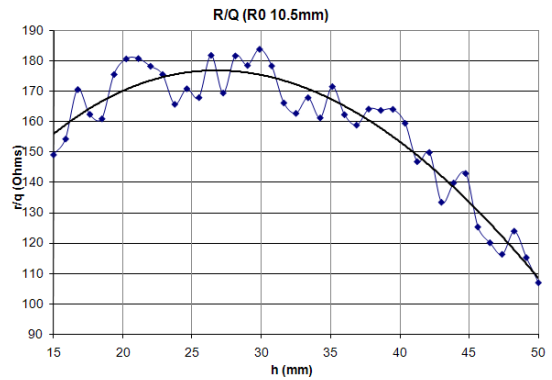
⁹full coarse parameter sweep

¹⁰only cavities with 1.3 GHz resonance

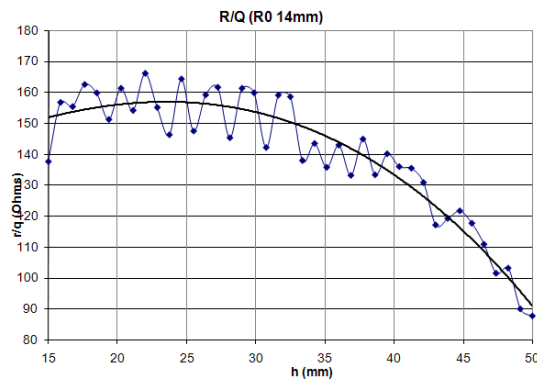
the cavity h . Observing the trend of R/Q vs R_0 the R/Q as expected decreases as R_0 (C) increases while decreasing L (h) to keep the frequency constant ($R/Q = \sqrt{L/C}$).



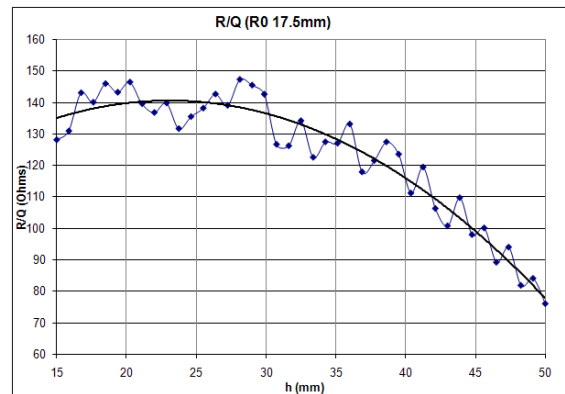
(a) $R_0=7$ mm



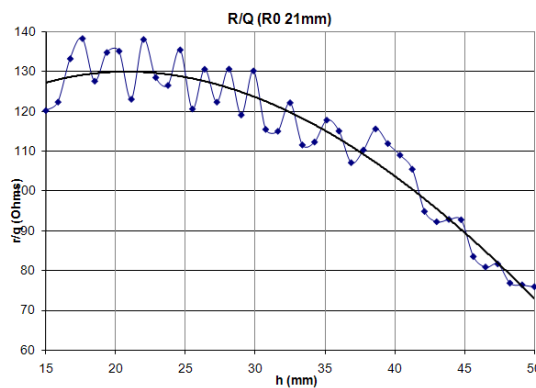
(b) $R_0=10.5$ mm



(c) $R_0=14$ mm



(d) $r_0=17.5$ mm



(e) $R_0=21$ mm

Figure 2.31: R/Q against h for coaxial cavities with frequency 1.3 GHz with varying R_0 . Increasing R_0 decreases maximum R/Q

Keeping R_0 to a minimum is therefore preferable but a decision can only be made after thermal calculations (see fig. 2.32).

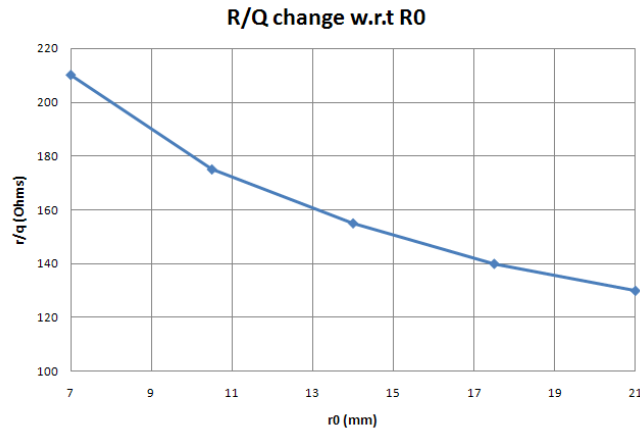


Figure 2.32: Coaxial cavity maximum R/Q (Ω) vs R_0 (mm)

2.11.2 HOM in Coaxial Cavity

The HOM of the coaxial cavity were calculated using the *JDM* solver in MWS with a mesh density of 30 mesh lines per wavelength and a frequency range of 0-7 GHz. The mode spectrum was split into two by assigning (x,y) symmetry of (magnetic,magnetic) and (magnetic,electric) in order to make use of the increase in accuracy and reduction in calculation time afforded by use of symmetry. The first set starting with the fundamental TM mode and the second the dipole mode. Only TM modes with electrical symmetry around $z = 0$ have been found.

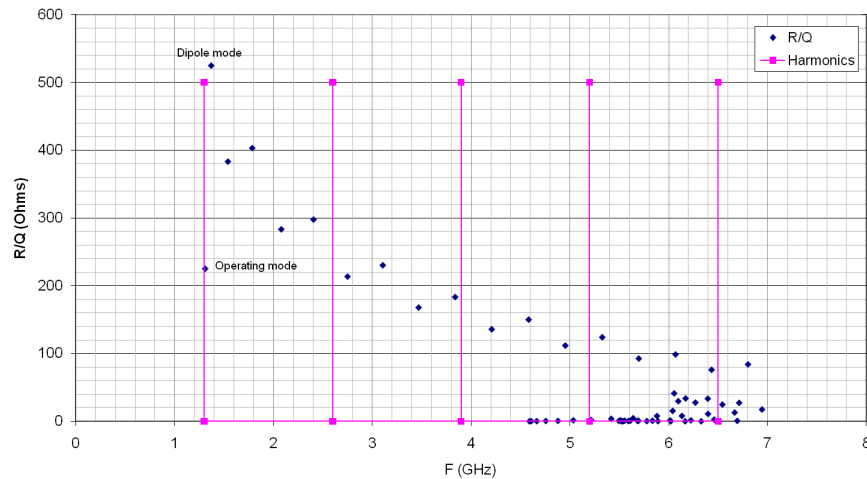


Figure 2.33: Coaxial cavity mode spectra up to 7 GHz showing R/Q and optimal R/Q and harmonics of 1.3 GHz. Of particular note is the mode very close to 1.3 GHz (a dipole mode).

The R/Q values were calculated for a single point at the electric field, maximum giving a worst case value.

Disturbingly the dipole mode is very close indeed to the fundamental TM mode, with the dipole frequency being 1.360 MHz and the target frequency being 1.3 GHz. The single point R/Q of this mode is in fact much higher than that of the fundamental mode at 500 Ω . If the beams could be guaranteed to be identical and modulated perfectly in phase this mode would not be a problem. Unfortunately in a practical klystron neither of these conditions are true.

If purely the R/Q is considered, coupling is strong for many modes. However as the beams are held in phase (by the input cavity modulating the beams in the correct mode) the coupling is highly dependent on the azimuthal index of the mode or beam current/modulation asymmetries.

For instance in the case of a two beam klystron:

- with ideal beams perfectly in phase dipole modes will not couple as the bunches will see fields in antiphase and so the induced voltages will cancel; however the quadrupole mode will couple as the bunches will see fields in phase (but of course not in a 4 or above beam klystron).

- with beams allowed to take on any phase (as in the input cavity with an incorrect mode excited); dipole modes will couple perfectly, as will the quadrupole mode (with reduced stored energy available for interaction).

As the beams are forced into phase by the input cavity the first case is most common. The next mode that can be excited with perfect beams is the $TM_{20,1,0}$.

If the external loading of the input cavity is high then the Q_0 curves of the fundamental and dipole modes could overlap allowing both to be excited. Care would be needed with the design of the input coupler to avoid this problem. The effects of the dipole mode on cavity-beam interaction are explored in chapter 5.

2.11.3 2^{nd} Harmonic Cavity

Table 2.18: Simulation range for coaxial 2^{nd} harmonic cavity

R_1	8-30 mm
h	7-30 mm

As the interaction structures (see chap. 6) and high efficiency operation call for a second harmonic cavity, a compatible cavity must be found with a fundamental resonant frequency of 2.6 GHz. Compatibility is ensured by retaining the r_{pcr} as constant locating the beams in the same place inside the cavity. This means that R_1 and h should be reduced to increase the frequency. This is acceptable as the frequency is only weakly dependant on r_{pcr} for $TM_{0,1,0}$ cavities.

The R/Q required need not be as high as for the other cavities so the optimum need not be as precise. The plot shows that a R/Q of over 120Ω is achievable for a wide range of cavities with dimensions well within the sensible limits (see fig. 2.34 and tab. 2.19). The use of such a cavity is preferable to trying to squeeze in re-entrant cavities between the beams.

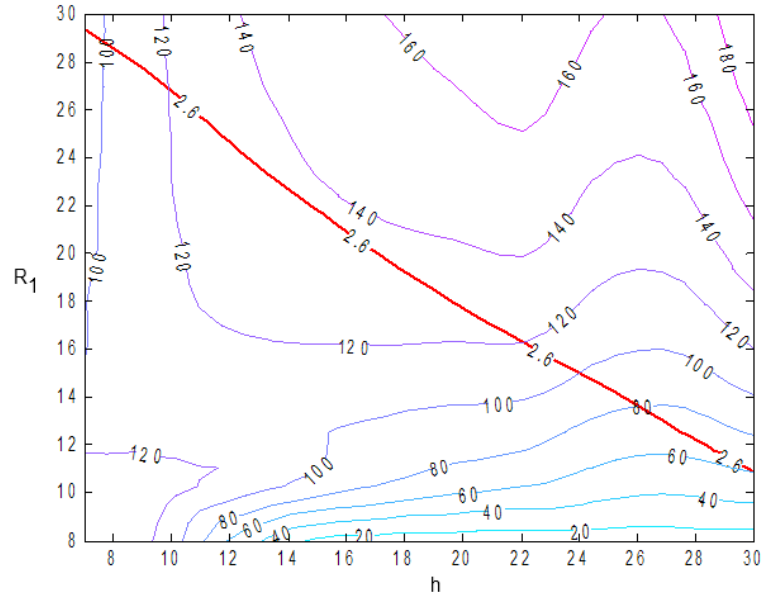


Figure 2.34: Coaxial 2nd harmonic cavity R/Q contours and 2.6 GHz contour

Table 2.19: 2nd harmonic cavity possible properties

Frequency	2.6 GHz
R/Q	>120 Ω
height (h)	10-22 mm
radius (R_1)	16-26 mm

2.12 Conclusions

Instructed by the preliminary conclusions (see sec. 2.8) the klystron parameters were reviewed and altered in order to avoid some potential problems such as output gap arcing. The two most promising cavity geometries were optimised again for the new parameters.

Due to the enforced r_{pcr} the recessed re-entrant cavity could not be re-tuned to support the target frequency and so is no longer a candidate. This cavity did however show promise and could prove useful for a different klystron specification.

The fundamental mode coaxial cavity's performance is degraded by the revised parameters, at 140 Ω to 210 Ω (dependant on R_0) as opposed to $\sim 250 \Omega$.

However it remains both compact and more suitable for an interaction structure than the HOM cavities prove to be.

From initial analysis of the HOM, the dipole mode is clearly problematic, with a mode separation of around 50 MHz. This small gap is combined with an R/Q two times larger. The effect of this mode needs to be evaluated and a solution proposed.

The coaxial cavity can also be used as a 2nd harmonic cavity, without any compatibility problems, with an R/Q of 120 Ω being easily achievable and suitable.

Table 2.20: Revised parameter cavity comparison

	1.3 GHz Resonance	
	Coaxial $TM_{0,1,0}$	Coaxial 2nd Harmonic $TM_{0,1,0}$
R/Q (Ω)	130-210	>120
Q	~8-9000	~8000
Outside radius r_1 (mm)	160	138.5
R_1 (mm)	35-40	20
r_{pcr} (mm)	118.5	118.5
r_t (mm)	5.83	5.83
R_0 (mm)	7-21	7
Height (h) (mm)	20-27.5	15
Damping	possible	

With the removal of the recessed re-entrant cavity, lacking further information about the HOM characteristics, the coaxial cavity remains a good option. The structure of the whispering gallery cavity would have changed little with the new parameters. Only the gap would have lengthened, so would not alter the cavity properties sufficiently for it to be preferable to the fundamental coaxial cavity. The higher R/Q of this cavity is more suitable to use in a klystron interaction structure and the compactness will ease the design of the focusing system.

Chapter 3

Dipole Mode

3.1 Introduction

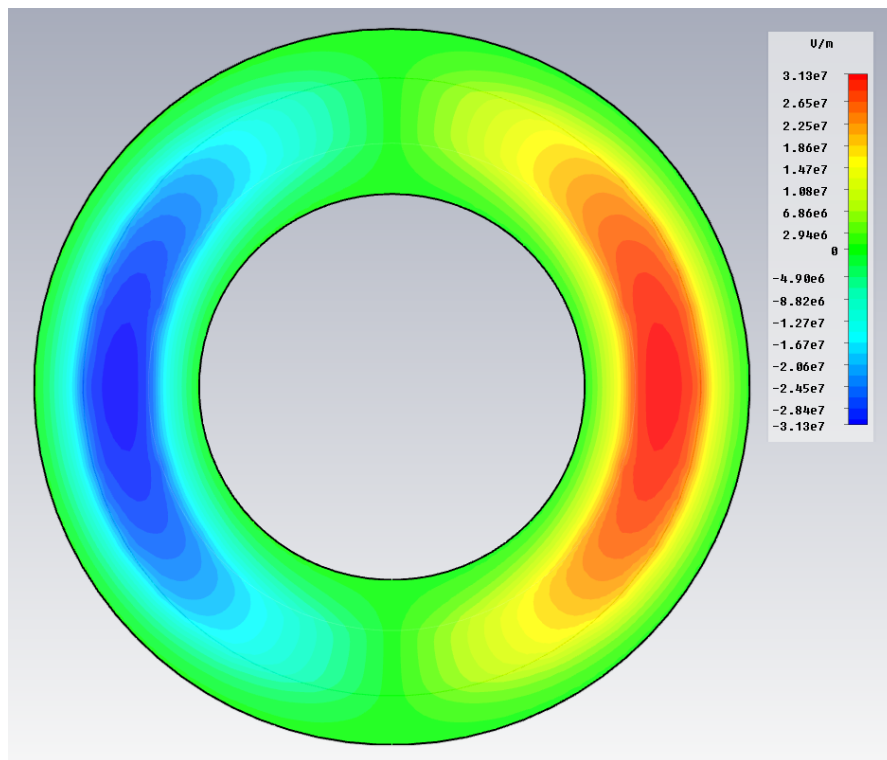


Figure 3.1: Dipole mode of the fundamental mode coaxial cavity

As noted in the cavity selection process (see sec. 2.11.2) the dipole mode of the coaxial cavity is close to the operating frequency of the klystron. It is unclear whether this will affect the stability of the MBK, an attempt to model

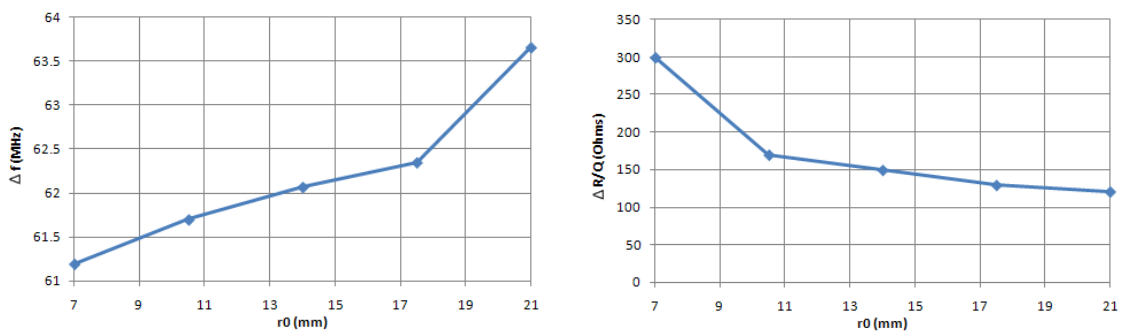
this is made in chapter 5.

During the design and construction of a similar coaxial cavity klystron by CPI[12] problems with the dipole mode were not experienced or expected. The dipole mode in their cavities geometry was calculated to be approximately 200 MHz from the operating frequency¹.

As the dipole mode in the coaxial cavity is only 60 MHz from the fundamental mode the reasons for this and remedial actions should be evaluated.

3.2 Effect of r_0 on Dipole Mode

In an effort to increase the separation r_0 is altered to broaden and narrow the beam ridge (see sec. 2.9.5). It was hoped this change in cross section would alter the separation. In addition to the change in r_0 , r_1 and h are re-tuned to maintain the correct frequency. Thus a small range of different cross sections are investigated.



(a) Dipole mode-fundamental mode frequency separation against r_0

(b) Separation of monopole and dipole mode R/Q against r_0

Figure 3.2: A shift of only around 63 MHz can be expected and the R/Q of the dipole mode remains above the monopole mode. (see fig. 2.32)

The HOM were calculated for the full range of r_0 in order to find the variation of the dipole mode. Fig. 3.2 shows the difference in frequency and the difference in R/Q between the dipole and fundamental mode with respect to r_0 . This shows that as r_0 is changed (and so h and r_1) the frequency difference bet-

¹A. Balkum, private communication

ween the dipole and fundamental mode changes a small amount from 61 MHz to 63.5 MHz (2.5 MHz) (see fig. 3.2). The R/Q of the dipole mode also remains over 100Ω above the R/Q of the fundamental mode.

This does suggest that for an optimal R/Q cavity within these constraints the dipole mode cannot be moved by trivially altering the cross section. Some form of perturbation is required to move the dipole mode to a frequency further away from the operating frequency.

3.3 Waveguide Theory

The coaxial cavity (see sec. 2.5) resembles a section of ridge waveguide bent into a circle. It could be modelled as a length of straight waveguide with a length of $2\pi r$ (where $r = pcr$) and with periodic boundaries at either end with a phase advance of $2\pi N$ (where N is a natural number)(see fig. 3.3). This analysis allows the use of well known waveguide theory to analyse the frequencies of the modes.

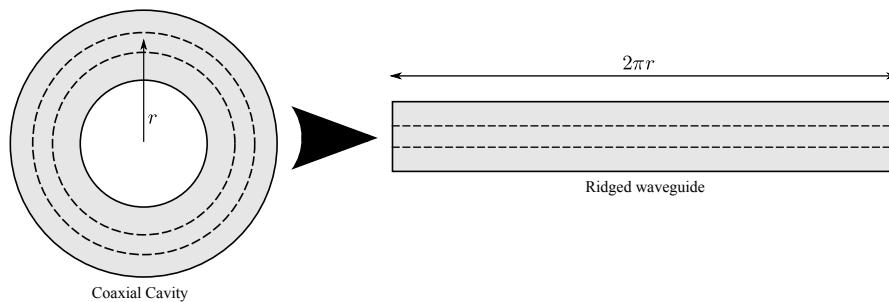


Figure 3.3: The conversion between coaxial cavity and ridged waveguide

The effect of the periodicity of the boundary conditions is to ensure that k_z takes on discrete values (β). Due to the shape of the dispersion curve (see fig. 3.4) if $\frac{d\omega}{dk_z}$ is large then the separation between the first two modes will increase disproportionately. This spacing can be calculated using wave guide theory.

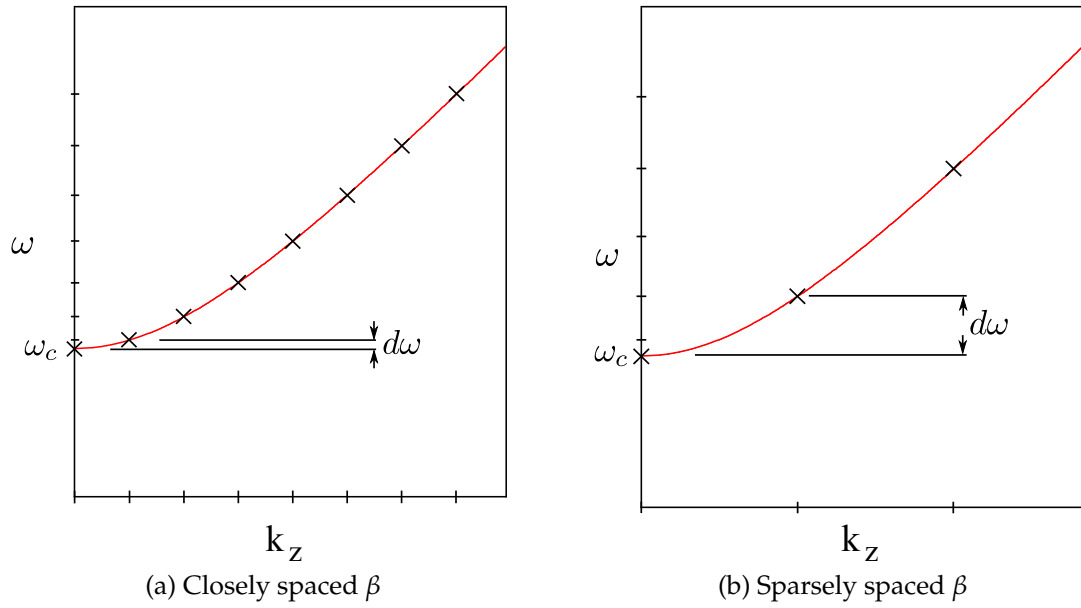


Figure 3.4: k_z, ω plot showing discrete wavelength solutions where the periodic boundary conditions are satisfied showing the mode spacing in ω increases with increasing β

The starting point is the dispersion relationship [13]:

$$k^2 = k_c^2 + k_z^2 \quad (3.1)$$

$$\left(\frac{\omega}{c}\right)^2 = k_c^2 + k_z^2 \quad (3.2)$$

The wavelength of a mode (and its wave number) can be defined as:

$$\lambda = \frac{2\pi r}{N} \quad (3.3)$$

$$k_z = \frac{2\pi}{\lambda} = \frac{2\pi N}{2\pi r} = \frac{N}{r} = \beta \quad (3.4)$$

where $N = 0, 1, 2, \dots, \infty$ and represents the number of wavelengths along the waveguide

Substitute k_z by $\frac{N}{r}$ in eq. 3.2 and multiply both sides by c^2 :

$$\omega^2 = \omega_c^2 + \left(\frac{cN}{r}\right)^2 \quad (3.5)$$

If both sides are divided by ω_c the ratio between ω and the cut off frequency ω_c is:

$$\frac{\omega}{\omega_c} = \sqrt{1 + \left(\frac{cN}{\omega_c r}\right)^2} \quad (3.6)$$

as this is an ideal lossless system $\frac{\omega_c}{c}$ can be replaced with k_c this leads to the result:

$$\frac{\omega}{\omega_c} = \sqrt{1 + \frac{N^2}{r^2 k_c^2}} \quad (3.7)$$

ω_c is both the cut off frequency and the fundamental mode. This can be calculated for waveguide or cavity using MWS or using a analytical ridge waveguide theory such as Hofer and Burton [65].

3.3.1 Theory Verification

In order to verify that this approach works a comparison is made with MWS simulations of the coaxial cavity investigated previously. Two modes initially chosen for this purpose are $TM_{x,1,0}$ and $TM_{x,2,0}$.

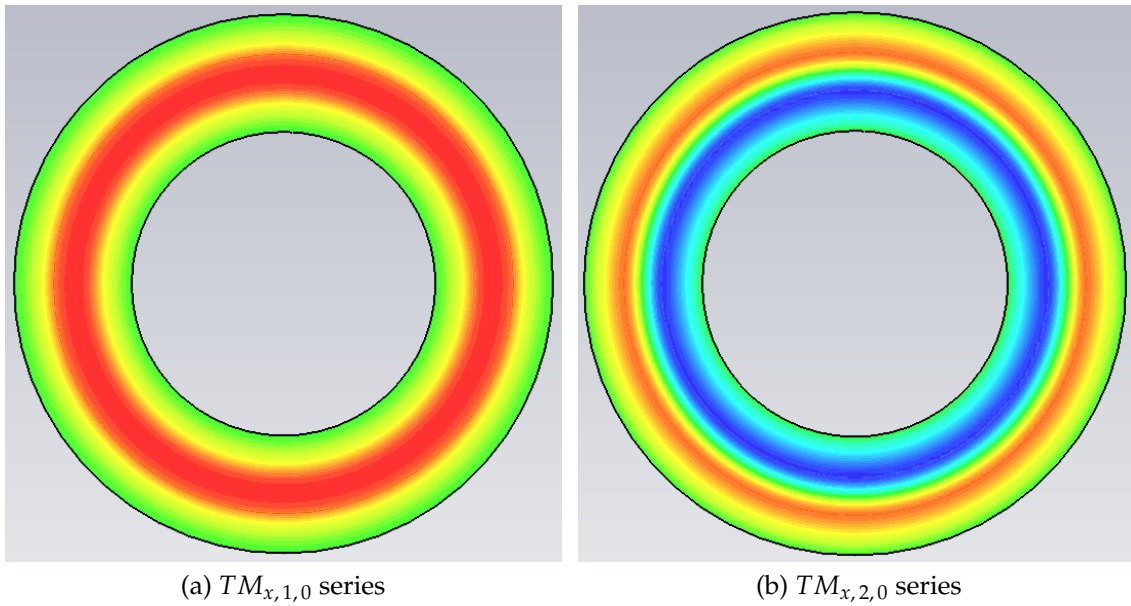


Figure 3.5: Plots of E_z for the two test series of modes in the coaxial cavity (in these cases $x = 0$)

Percentage errors between the MWS frequencies and the theory predictions are shown (see fig. 3.6).

The first mode error is 0% as the frequency is taken directly from MWS and used as the seed ω_c for the prediction equations. The accuracy of the prediction decreases as the mode number (number of wavelengths around the cavity) increases. The curvature of the cavity could be affecting the results to a greater extent in this region.

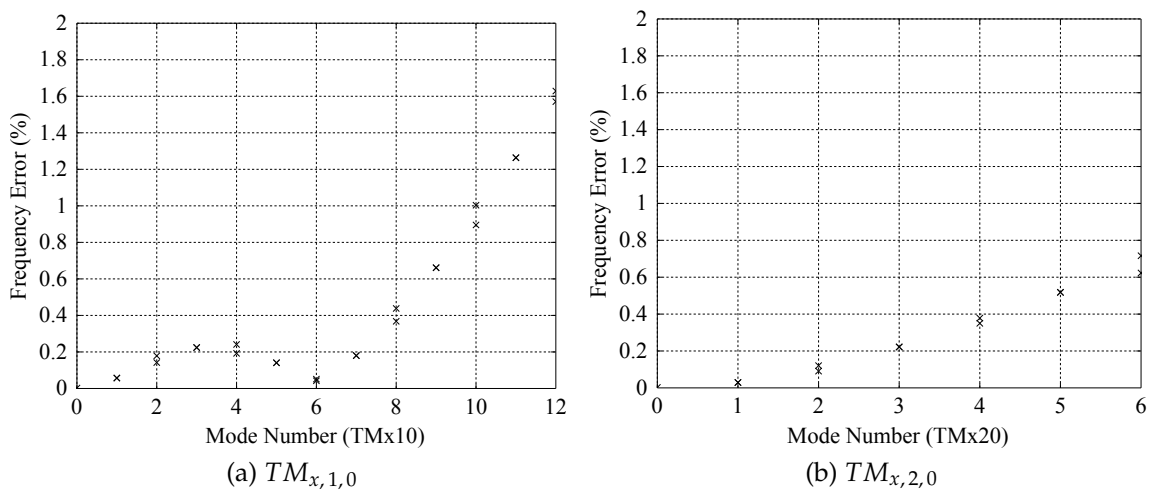


Figure 3.6: Percentage deviation of the frequency prediction from the MWS results

There is no reason that this approach should not be applied to other kinds of modes as long as their cut off frequencies are supplied. With this in mind the theory was tested for the TE_{x11} modes (see fig. 3.7), this mode is in fact found in two positions, both with an associated series of modes.

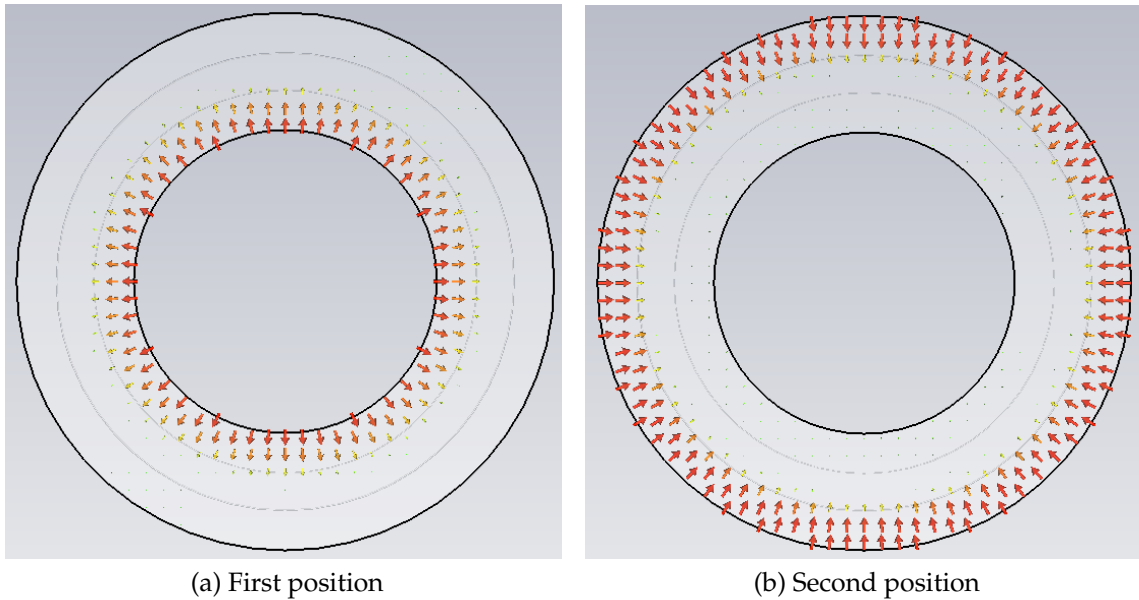


Figure 3.7: Plots of E for the two positions of the TE_{x11} series of modes ($x = 0$ for these plots)

In the first instance there is a high degree of error between the predicted and actual results for both positions of up to 12% (see fig. 3.8).

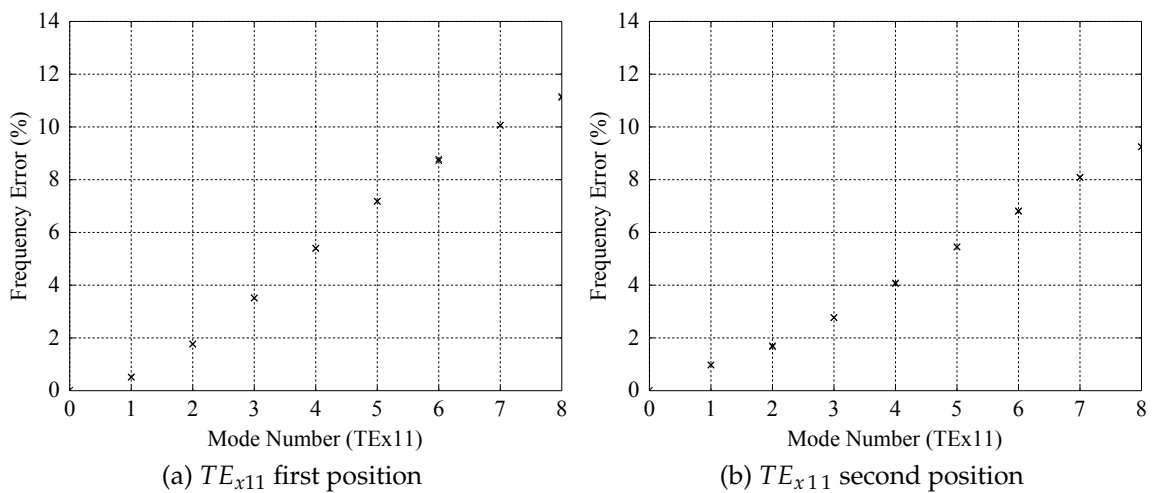


Figure 3.8: Percentage deviation of the frequency prediction from the MWS results

This is due to the fact that the radius of each of these modes is noticeably different. The $TM_{x,x,0}$ modes are well centred in the waveguide; the TE_{x11} modes with their dual positions are not.

As the cavity effectively has re-entrant nose cones it is possible for modes to be trapped in the inner and outer sections. A careful choice of R should be made using the mode's centre not the cavity's.

Instead of using the centre point of the cavity as in fig. 3.3, the radius can be taken at the lower and upper quarter of the cavity for the first and second positions respectively. Using this correction the frequency error is brought down to below 1% (see fig. 3.9).

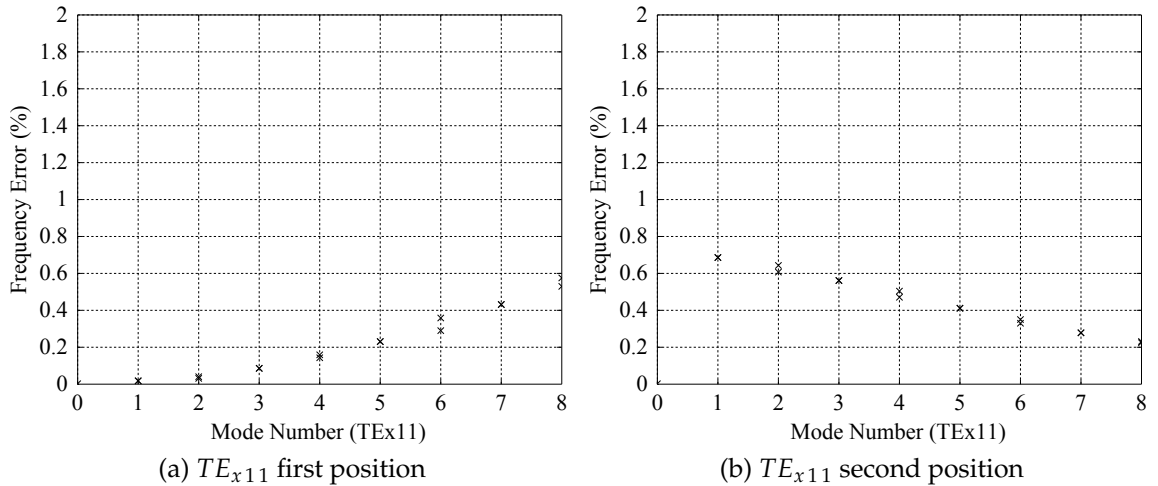


Figure 3.9: Percentage deviation of the frequency prediction from the MWS results

3.3.2 Results

Using eq. 3.7 the ratios of the dipole and quadrupole modes can be plotted against the r_{pcr} of the cavity (see fig. 3.10). The radius of the cavity used by CPI is approximated using knowledge of the number of beams and the frequency separation observed. Fig. 3.10 clearly shows that the 6 beam configuration which allows CPI to use a smaller cavity also leads to a large separation between the monopole and dipole modes. A 20 beam configuration mandates a large cavity and, from this analysis, a smaller separation between monopole mode and di-

pole. Using the full power CLIC specification is likely to increase the size of this cavity further, if only due to thermal constraints, and so decrease the mode spacing further.

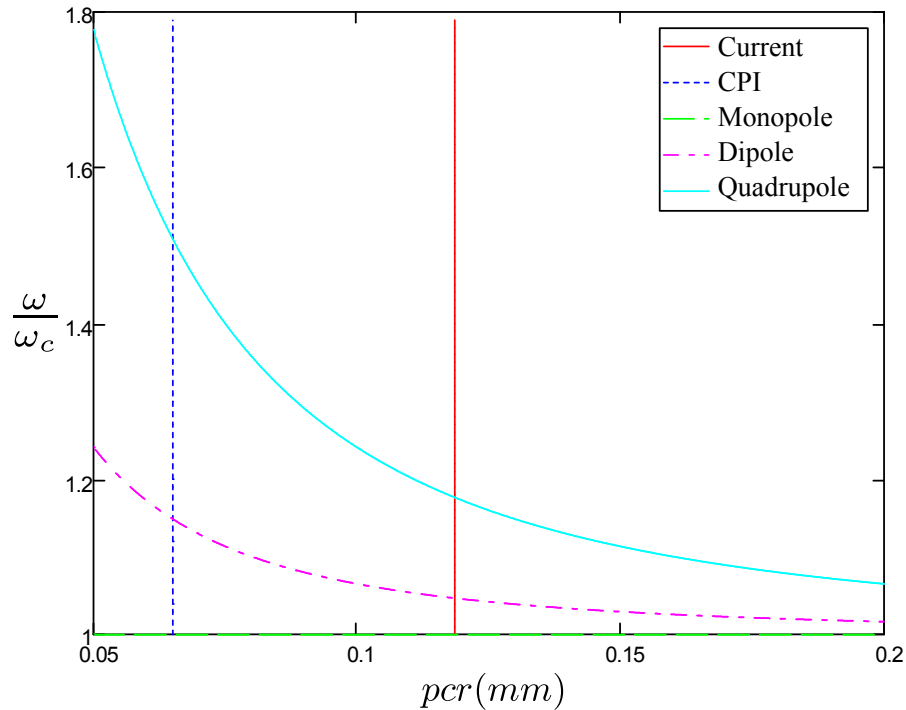


Figure 3.10: Ratio of monopole mode to dipole and quadrupole mode for various cavity sizes showing the CPI design[12] and the current design

Interestingly eq. 3.7 does not depend on the cross section of the waveguide, only the cut-off frequency which in this case is fixed at 1.3 GHz. This property suggests that very little can be done by altering the cross section. The attempt made previously (see fig. 3.2) found that a change in r_0 (and a re-tuned cavity) could only effect a ~ 2.5 MHz change in dipole frequency shift (see fig. 3.2). This theory has little to say about the R/Q which is strongly linked to the cross section as evidenced in the associated R/Q plot (see fig. 3.2).

The separation could be increased by altering the width of the waveguide, but this is constrained by the required frequency of the cavity.

3.4 Shifting Gap

3.4.1 Dipole Mode Suppression

The dipole mode remains a potential problem at ≈ 60 MHz from the operating mode. The waveguide theory approach (see sec. 3.3) suggests that the cross section of the cavity is unimportant and that the ratio of the modes is constant for constant radius. The only simple solution is to decrease the radius to increase the spacing (see fig. 3.10). As the klystron has been optimised for efficiency and small cavity radius, this is not possible without alteration to the design.

There are a number of options:

- Reduce the number of beams
- Position the beams in multiple rings
- Perturb the cavity to separate the modes
- Selectively damp the dipole mode

Reducing the number of beams would increase the perveance and so lower the potential efficiency of the tube. As efficiency is a major design goal this is undesirable.

Positioning the beams in multiple rings could solve the problem by allowing a smaller cavity to be used, but the beams will no longer be identical, further complicating the design process. As the similarity of the beams is a distinct advantage in interaction structure design, this is also undesirable.

Selectively damping the dipole mode is problematic as the electric and magnetic fields only exist in the same locations in the dipole mode and the operating modes. One possibility is to couple the modes to auxiliary cavities which are tuned to perturb the dipole mode and can be filled with lossy material.

Waveguide theory may again provide an insight.

3.4.2 Theory

Investigating the cavity as if it were waveguide is again useful. The operating mode and dipole modes can be represented as TE_{01} mode in rectangular waveguide. [13] contains an equivalent circuit to represent this mode (see fig. 3.11).

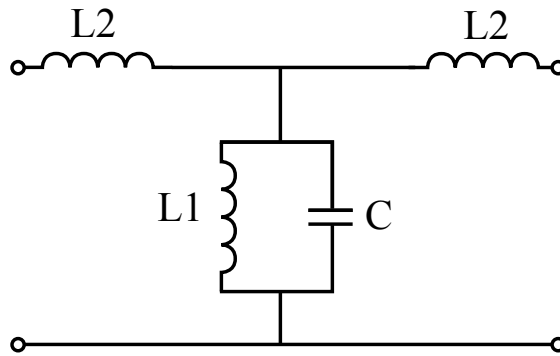


Figure 3.11: Equivalent circuit representing the $TE_{0,1}$ in rectangular waveguide[13]

From Section 3.3 eq. 3.6 can be taken as a starting point:

$$\frac{\omega}{\omega_c} = \sqrt{1 + \frac{c^2 N^2}{\omega_c^2 r^2}} \quad (3.8)$$

it has been shown by Carter [13] that for this arrangement (see fig. 3.11) there is the relation:

$$\frac{\omega_c^2}{c^2} = \frac{L_2}{L_1} \quad (3.9)$$

so substituting this into eq. 3.8:

$$\frac{\omega}{\omega_c} = \sqrt{1 + \frac{N^2 L_1}{r^2 L_2}} \quad (3.10)$$

In this case it is the ratio of the inductances (along with the radius r which shall be kept fixed) which determines the mode ratio. Either L_1 can be increased or L_2 decreased to increase the mode separation.

The components represent the fields in the following ways:

$$\begin{aligned} L_1 &\rightarrow B_r \\ L_2 &\rightarrow B_\theta \\ C &\rightarrow E_z \end{aligned}$$

Altering B_θ (L_2) is far easier in this case and has the advantage of the possible solutions being rotationally symmetrical. The proposed solution is to add a slice of vacuum in the centre of the cavity (see fig. 3.12). This will act in one sense to stop the cavity behaving like a piece of ridged waveguide. It also breaks the current path along the inside face of the cavity perturbing the magnetic field and so the inductor which represents it. Such a change could be expected to lower this inductance as desired.

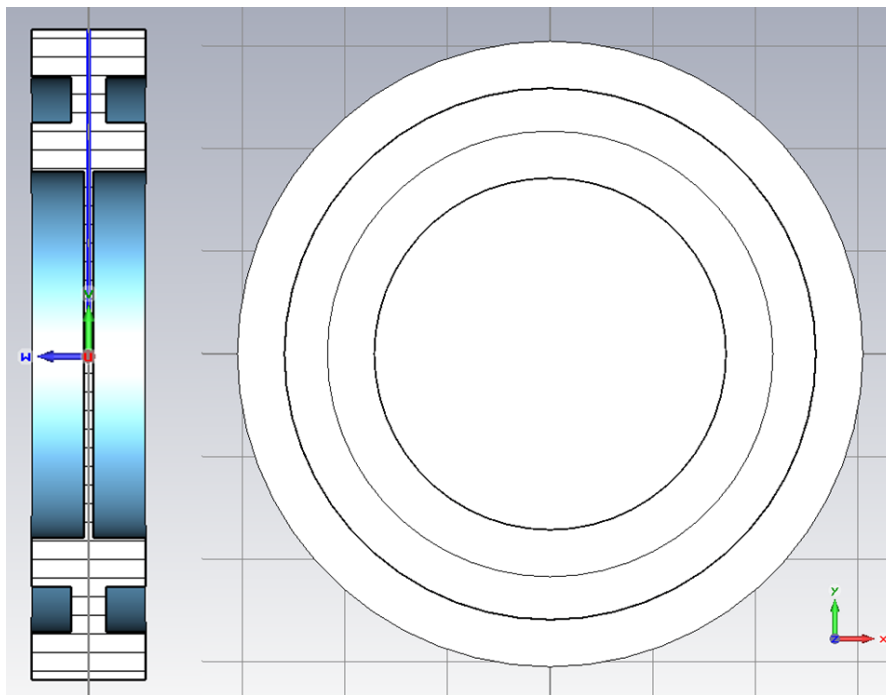


Figure 3.12: Proposed dipole shifting design shown in “solid” vacuum view

3.4.3 Results

For a range of shifting gaps it can be shown that the optimum value is 2.5 mm. As the gap increases the frequency of the dipole mode increases.

In addition the frequency of the quadrupole mode decreases with increasing gap. The optimum shifting gap is therefore 2.5 mm where the two modes cross (see fig. 3.13). This yields a frequency shift between the two modes of ≈ 125 MHz (see table. 3.1).

Table 3.1: Modes of interest in the dipole shifted cavity

Mode	Frequency
Operating $TM_{0,2,0}$	1.305 GHz
Dipole $TM_{1,2,0}$	1.432 GHz

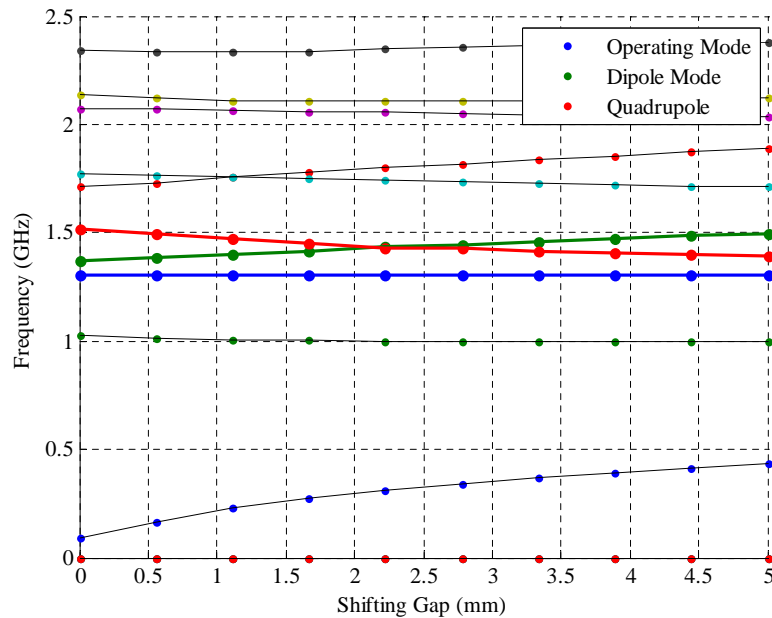
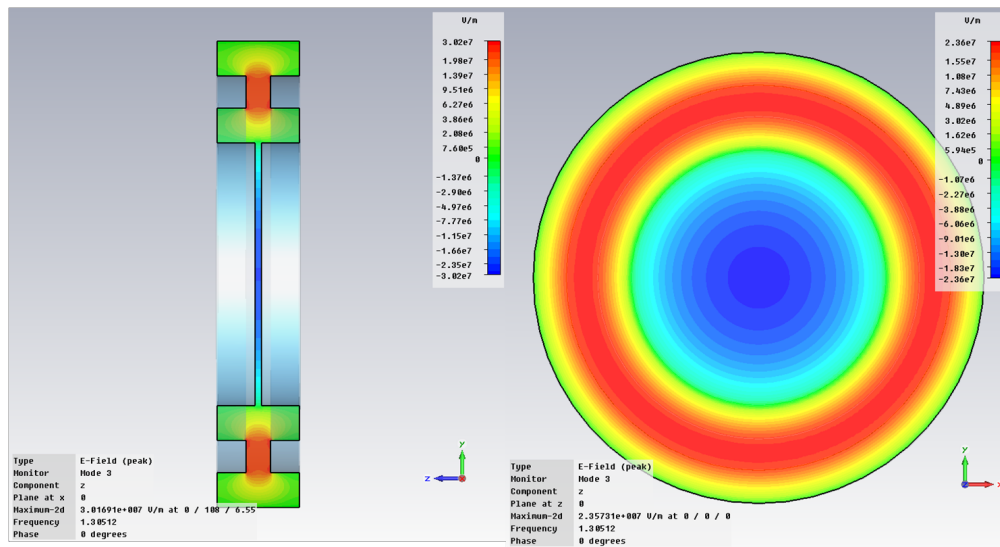


Figure 3.13: Frequency against the size of the shifting gap in mm. As the gap increases from 0 the dipole mode moves away until it crosses the descending quadrupole mode. It is at this point that the optimum gap lies.

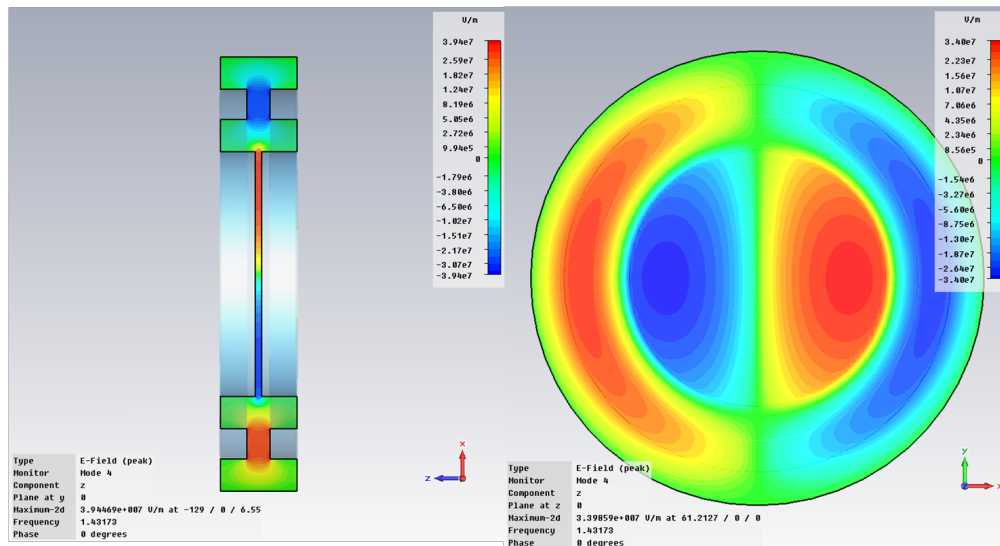
The two modes of interest were previously HOM in the standard cavity, with the operating mode now the $TM_{0,2,0}$ and the dipole $TM_{1,2,0}$. There is E_z field (of a similar strength to that at the beam radius) in the centre of the cavity which did not previously exist, which is now present both of the new operating mode and new dipole mode.

This situation occurs because the gap in the inner conductor now behaves as a thin pillbox cavity of a radius equal to the co-axial inner radius. This couples

to the previous coaxial modes.



(a) Operating mode $TM_{0,2,0}$



(b) Dipole mode $TM_{1,2,0}$

Figure 3.14: E_z plots of new modes which interact with the beam, now split by ≈ 125 MHz

This coupled pillbox cavity introduces new LOMs (Lower Order Modes) as it behaves similarly to an independent pillbox. However their field strength at the beam radius is very much reduced and they are far from the operating frequency at 0.3 GHz and 1 GHz for the fundamental and dipole modes respectively. There are of course higher order modes than these with similar properties, but the low field in the beam region should ensure that even with a comparable frequency to that of the tube there should be little interaction. However if

indeed these can be excited by the beam damping is non trivial.

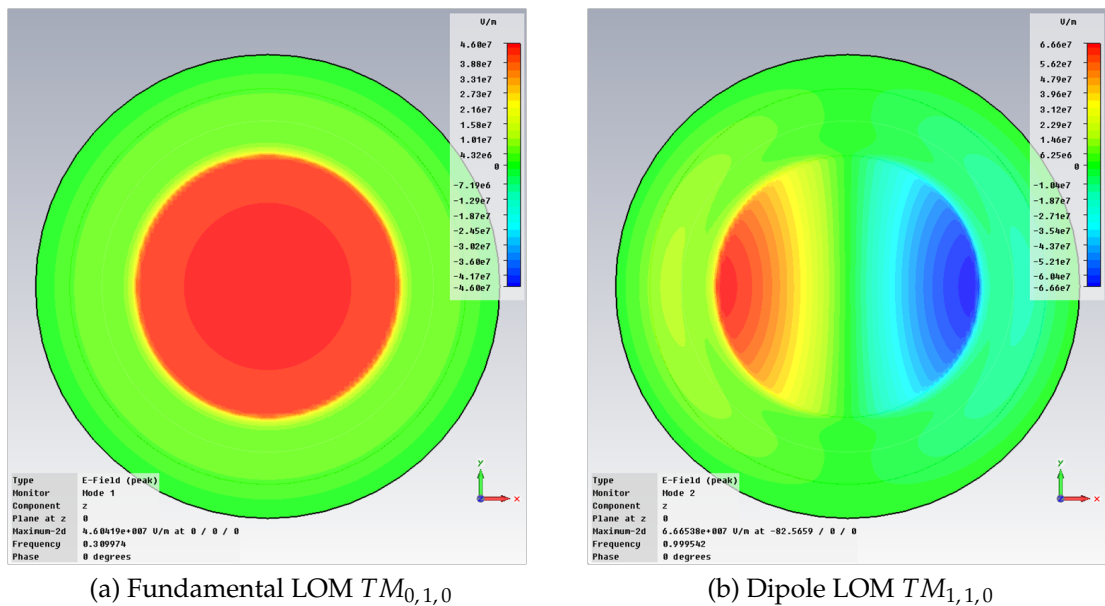


Figure 3.15: The two new LOM (Lower Order Modes) which may be hard to damp. However the E-field is low in the beam region and their frequency is far from the operating frequency

3.4.4 Central Post

The mechanical stability of the new cavity design could become a problem in a full tube. If the shifting gap could be terminated centrally the LOMs (see fig. 3.15) could be removed and the cavity stiffened (see fig. 3.16).

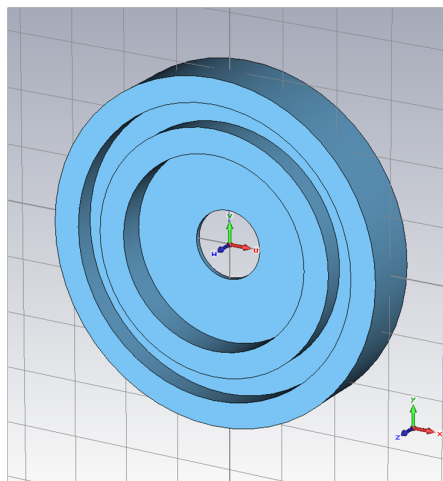


Figure 3.16: Shifting gap design with a central post shown in “solid” vacuum view.

For even a small radius of post this moves the new operating mode $TM_{0,2,0}$ and new dipole mode $TM_{1,2,0}$ together (see fig. 3.17). However if mechanical stability is a problem a very small central post (of the order of 1 mm radius) could be used for a small loss of dipole mode separation.

When a range of post sizes are plotted, and the modes identified, the transition of the modes between the shifting gap design and the standard design can be seen (see fig. 3.17).

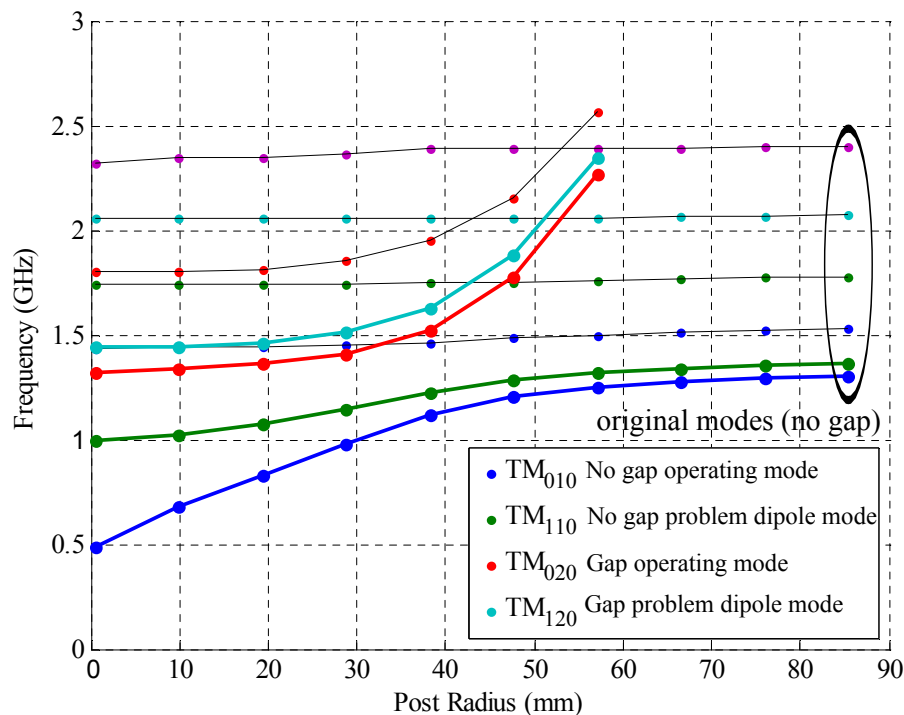


Figure 3.17: The transition between no central post (post radius = 0) and no shifting gap (post radius = r_1) (the termination of lines are due to the frequency range of the simulation). The original operating mode and dipole ($TM_{0,1,0}$, $TM_{1,1,0}$) can be seen to transition into the new LOMs and the new operating mode and dipole ($TM_{0,2,0}$, $TM_{1,2,0}$) transition into a HOM.

As the gap is introduced to the basic coaxial cavity, the original operating mode $TM_{0,1,0}$ transitions into the fundamental LOM $TM_{0,1,0}$ in the shifting gap and its associated dipole mode into the dipole LOM $TM_{1,1,0}$ (see fig. 3.15). Two higher order modes the $TM_{0,2,0}$ and the $TM_{1,2,0}$ reduce in frequency from a very high order mode to become the new operating modes 3.14.

3.5 Conclusions

This chapter attempts to model the coaxial cavity dipole mode using simple waveguide theory. The cavity can be accurately represented as ridged waveguide which leads to the conclusion that the dipole mode ratio is fixed almost regardless of cross section (if ω_c is constant). A relationship between the fundamental and dipole modes can be derived from this theory. This shows good agreement within 2% to MWS for both TM and TE modes of several orders. The R/Q cannot be predicted using this method as it is highly sensitive to cross section.

The coaxial cavity used by CPI has a much smaller radius giving them a naturally larger dipole mode separation. However, due to its size, this design will not scale to higher powers.

A simple equivalent circuit model shows that the dipole mode separation can be altered by lowering the inductance (L_2) at the centre of the cavity by perturbing B_θ . The addition of a vacuum gap across the centre of the cavity to achieve this has also changed the operating mode from $TM_{0,1,0}$ to $TM_{0,2,0}$.

The optimum value of this, with respect to dipole separation, is 2.5 mm. This moves the dipole mode (now $TM_{1,2,0}$) from 1.35 GHz to 1.43 GHz increasing the split from ≈ 50 MHz to a much more comfortable ≈ 125 MHz.

Both the dipole and operating mode transition between $TM_{x,1,0}$ and $TM_{x,2,0}$ as the extent of the cut reaches the centre of the cavity (a full slot). This also leads to the addition of lower order modes ($TM_{0,1,0}$, $TM_{1,1,0}$...) in the centre of the cavity. As the beams do not pass through this region they should not be excited. A simple solution could be to introduce a small central post with a radius of the order of 1 mm. This would add rigidity while only perturbing these LOM.

This appears to be a suitable solution to the problem of the dipole mode for the intermediate cavities. The input and output couplers may require the use of another solution.

Chapter 4

Method Of Moments

4.1 Introduction

Stability is key to the operation of any klystron. There is great potential for instability with a design which employs 20 beams due to:

- Monotron oscillations (see sec. 1.3.5.3)
- Beam asymmetries
- Unwanted modes

Accurate calculation of the modes is critical to identifying and rectifying any problems with the interaction structure. These higher order modes (HOMs) can be found by using existing techniques such as those employed in CST MWS¹ and MAFIA². However obtaining accurate information about the higher order modes using these codes can be problematic and very time consuming. The singularity caused by the sharp corner limits the ability of FDTD³ or FE⁴ methods to accurately model the otherwise simple geometries of, for instance a re-entrant cavity. Increased mesh around the sharp corner (and often by necessity the whole geometry) can help to alleviate the effects of this singularity but

¹CST MicroWave Studio <http://www.cst.com/Content/Products/MWS/Overview.aspx>

²CST MAFIA <http://www.cst.com/Content/Products/MAFIA/Overview.aspx>

³Finite Difference Time Domain

⁴Finite Element

cannot solve the problem entirely. A fundamental problem when calculating high frequencies, even in simple geometries, is that as the frequency increases the mesh density must also increase in order to ensure that the mesh (and its associated equations) can fully represent the fields which are being calculated. The increase in mesh required to calculate high frequencies and model the sharp corner accurately, appreciably increases calculation time.

An alternative is to represent the electric and magnetic fields in each region of the cavity by infinite series of basis functions, which are then matched by the imposition of a continuity condition at the interface ($r = r_1$). The method of moments has been shown to be a very accurate and rapid method for calculating the fundamental *TM* modes of a re-entrant resonant cavity by Carter et al. [66, 67]. This work is based on a refinement of work by Hansen Chu [68] and Taylor [69] and is closely related to a method used by Schwinger and Saxon [70] to calculate admittances of discontinuities in waveguides. The fundamental physics and equations behind the problem clearly suggest that such a method could be used to find higher order modes of the same geometry.

The definition of *TM* modes does not hold in the transition between pillbox cavity and reentrant cavity. In this work modes which closely resemble the true *TM* modes will be described as *TM* and likewise for those which resemble *TE*.

4.1.1 Method of Moments

A simple singly re-entrant cavity is divided into two regions. The electric and magnetic fields are represented by infinite series of basis functions which satisfy Maxwell's equations and the boundary conditions for each region. If the infinite series of basis functions are used, the boundary conditions at the interface between the regions can be matched exactly. For a real world calculation the series must be finite, which means the *E* and *H* fields can no longer be matched at every point along the interface. However the continuity of the Poynting flux is maintained by the method used which means that the stored energy in

a loss-less cavity is constant (see sec. 4.2.4). This has been shown to be the appropriate boundary condition when performing these calculations for resonant cavities [66].

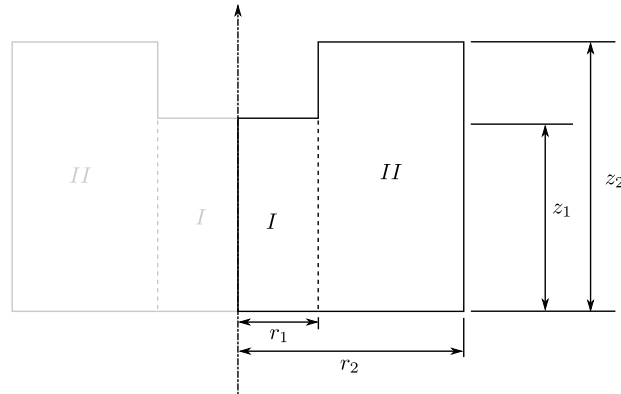


Figure 4.1: Representation of a re-entrant cavity used for the method of moments. Note the division into region *I* and region *II*.

It can be shown that the method can be extended to three (see fig. 4.2) regions but for simplicity two regions are used during the derivations following as the techniques and formulae can be manipulated directly to the three region form. With three regions all of the cavity types suggested can be modelled, even the coaxial cavities which can be modelled with a change of boundary condition at the inner edge of region *II*.

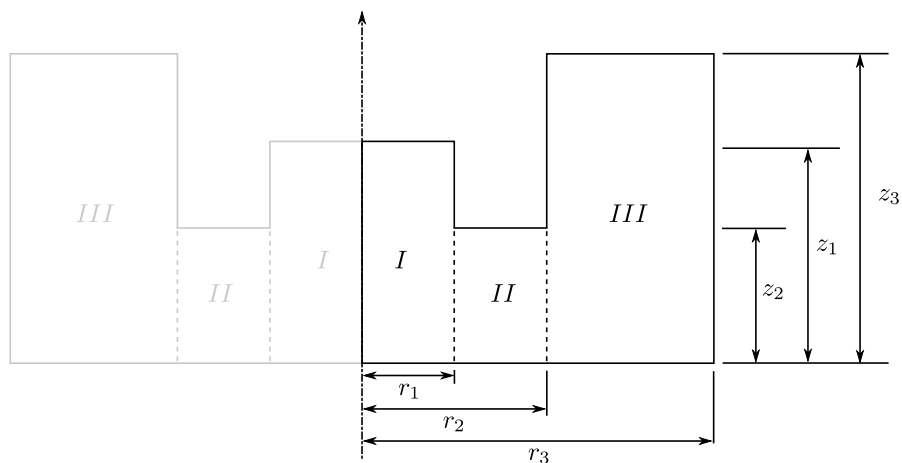


Figure 4.2: Three regions configuration

4.2 *TM* Mode Derivation

This method as described in section 4.1.1 represents the fields in each region as a series of basis functions. These functions must satisfy Maxwell's equations as well as the boundary conditions and can be represented in the form of a Fourier series.

One way of solving this problem has been described by Carter et al. [66]. The end result being a set of four matrix equations, one pair which link E and H in the same region and another pair which link E and H between regions.

Unlike Carter et al. [66] the following derivation includes the terms which represent the azimuthal variation of the fields for reasons which will become apparent in section 4.3. The original method was only developed for use with monopole modes but is suitable for use with other *TM*-like modes with no azimuthal variation.

4.2.1 Field Component Equations

This derivation differs slightly from that of Carter et al. [66] who made use of the fact that the fields of a cylindrically symmetrical *TM* mode in a resonator can be derived from a scalar electric Hertzian potential function which satisfies the equation [71]:

$$\nabla^2\phi + k^2\phi = 0 \quad (4.1)$$

This derivation is based on Eq. 4.2 as it also lends its self to computation of the *TE* mode basis functions. Starting with the transverse part of the Laplacian expressed in circular cylindrical coordinates from Ramo, Whinnery, and Duzer [72] which is clearly related to Eq. 4.1.

$$\nabla_t^2 E_z = -\gamma^2 E_z \quad (4.2)$$

where $\gamma = k_c = \sqrt{k^2 - k_z^2}$; $k_z = \omega/c$ in free space; $k =$ wave number; $E_z =$ electric field maximum in the z direction.

Eq. 4.2 can be expanded to find:

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial E_z}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 E_z}{\partial \theta^2} = -\gamma^2 E_z \quad (4.3)$$

which has a solution in terms of Bessel functions [73] of the form:

$$E_{z \text{ freespace}}(r, \theta) \sim \left\{ \begin{array}{l} AJ_i(\gamma r) + BY_i(\gamma r) \\ AI_i(\gamma r) + BK_i(\gamma r) \end{array} \right\} \cos(i\theta) e^{jk_z z} \quad (4.4)$$

where $i =$ azimuthal variation index; A and $B =$ constants;

Note the $\cos(i\theta)$ term to represent azimuthal variation (a $\sin(i\theta)$ variation is of course equally valid) and the order of the Bessel functions, also i (for instance in J_i). i is the index of the azimuthal variation, this is the key part of the equations which was neglected by Carter et. al. due to their working assumptions. With this term in place the equations are able to fully represent TM modes with azimuthal variation.

From Maxwell's equations the electric and magnetic field components can be defined as such [72]:

$$E_r = \frac{-j}{\gamma^2} \left[k_z \cdot \frac{\partial}{\partial r} \cdot E_z(r, \theta) + \frac{\omega \cdot \mu}{r} \cdot \left(\frac{\partial}{\partial \theta} \cdot H_z(r, \theta) \right) \right] \quad (4.5)$$

$$E_\theta = \frac{j}{\gamma^2} \left[\frac{-k_z}{r} \cdot \frac{\partial}{\partial \theta} \cdot E_z(r, \theta) + \omega \cdot \mu \cdot \left(\frac{\partial}{\partial r} \cdot H_z(r, \theta) \right) \right] \quad (4.6)$$

$$H_r = \frac{j}{\gamma^2} \left[\frac{\omega \cdot \epsilon}{r} \cdot \frac{\partial}{\partial \theta} \cdot E_z(r, \theta) - k_z \cdot \left(\frac{\partial}{\partial r} \cdot H_z(r, \theta) \right) \right] \quad (4.7)$$

$$H_\theta = \frac{-j}{\gamma^2} \left[\omega \cdot \epsilon \cdot \frac{\partial}{\partial r} \cdot E_z(r, \theta) + \frac{k_z}{r} \cdot \left(\frac{\partial}{\partial \theta} \cdot H_z(r, \theta) \right) \right] \quad (4.8)$$

where $\epsilon =$ permittivity of free space; $\mu =$ permeability of free space; $c =$ speed of light and $\omega =$ angular frequency

To characterise the mode the individual field components need to be found, which for a *TM* mode exclude H_z :

$$\overbrace{E_z \quad E_r \quad E_\theta \quad H_r \quad H_\theta}^{TM}$$

This allows the redefinition of Eq. 4.5-4.8, using the definition of $H_z = 0$, to find all of the field components in terms of E_z .

$$H_r = \frac{j\omega\epsilon}{r\gamma^2} \frac{\partial E_z}{\partial \theta} \tag{4.9}$$

$$H_\theta = \frac{-j\omega\epsilon}{\gamma^2} \frac{\partial E_z}{\partial r} \tag{4.10}$$

$$E_r = \frac{-1}{\gamma^2} \frac{\partial}{\partial r} (ik_z E_z) \tag{4.11}$$

$$E_\theta = \frac{-1}{\gamma^2 r} \frac{\partial}{\partial \theta} (ik_z E_z) \tag{4.12}$$

4.2.2 Application of Boundary Conditions

These field components are valid for a cylindrical waveguide, where as a cavity of course has a finite extent. In the case of region *I* & *II* there are conducting boundaries at $z = 0$ & $z = z_2$. With the knowledge that the z variation must be sinusoidal, it can be inferred from the boundary conditions that there must not be tangential electric field or perpendicular magnetic field at a conductor (see table. 4.1).

Table 4.1: The boundary conditions of *TM* modes

	<i>TM</i>	
	@ z_0 & z_2	z variation
E_r	$\neq 0$	$\cos(n\beta z)$
E_θ	0	$\sin(n\beta z)$
E_z	0	$\sin(n\beta z)$
H_r	$\neq 0$	$\cos(n\beta z)$
H_θ	$\neq 0$	$\cos(n\beta z)$
H_z	-	-
$n = 0, 1 \dots \infty$		

where n = basis function index and β is fixed by the length (see bellow)

n in Tab. 4.1 refers to the mode number in the z direction in a pillbox. This index takes on a special importance when the fields in the two regions are matched at the boundary.

So from eq. 4.4:

$$E_{z \text{ pillbox}}(r, \theta, z) = \left\{ \begin{array}{l} AJ_i(\gamma_n r) + BY_i(\gamma_n r) \\ AI_i(\gamma_n r) + BK_i(\gamma_n r) \end{array} \right\} \cos(i\theta) \cdot \cos(n\beta z) \quad (4.13)$$

where i = azimuthal variation index; γ_n becomes $\sqrt{k^2 - n^2\beta^2}$ as $k_z = n\beta$ and β is π/z_1 and π/z_2 in region I and II respectively as $n\beta z_x = n\pi$.

In order for the functions to represent the fields inside a cavity the boundary conditions must be applied to the solutions (see eq. 4.13). For these purposes E_z is defined as:

$$E_{z \text{ pillbox}}(r, \theta, z) = R \cdot \cos(i\theta) \cdot \cos(n\beta z) \quad (4.14)$$

where:

$$R = (AJ_i(\gamma_n r) + BY_i(\gamma_n r)) \quad (4.15)$$

There are two boundary conditions; Eq. 4.16 that E_z must be symmetrical and the 1st derivative must be continuous over $r = 0$; Eq. 4.17 that E_z must be zero at the outer conductor .

$$\left\{ \begin{array}{l} \frac{\partial E_z}{\partial r} = 0 \quad \text{for } i = 0 \\ E_z = 0 \quad \text{for } i \neq 0 \end{array} \right\} @ r = 0 \quad (4.16)$$

$$E_z = 0 @ r = r_2 \quad (4.17)$$

The two forms of R which meet these boundary conditions are:

Region I (4.16):

$$R^I = J_i(\gamma_n r) \quad (4.18)$$

Region II (4.17):

$$R^{II} = (Y_i(\gamma_n r_2) J_i(\gamma_n r) - J_i(\gamma_n r_2) Y_i(\gamma_n r)) \quad (4.19)$$

With the fields defined in each region the relationship between the E field and the H field in both region I and region II can be found.

In order to allow the fields to be matched at the boundary it is vital that the modes in each section can be mixed. This is achieved using the following infinite summation:

$$E_z(r, \theta, z) = \sum_{n=0}^{\infty} a_n \cdot R_n(\gamma_n r) \cdot \cos(i\theta) \cdot \cos(n\beta z) \quad (4.20)$$

where n is the index of the mode in the region as defined in 4.1 and a_n is the amplitude of a term in the series.

When modeling a pillbox the amplitudes of the terms with an index other than that of the mode represented will be 0. If a re-entrant cavity is calculated the amplitude of the terms will be those which best match the field across the interface (see sec. 4.2.4).

4.2.3 Convert Between E and H (\mathbf{G}^I and \mathbf{G}^{II})

The relationship between the E field and the H field is defined in the following terms:

$$\mathbf{h}_\theta = \mathbf{G}(k) \mathbf{e}_z \quad (4.21)$$

where \mathbf{e}_z is a vector of all e_{zn} s (where e_{zn} is related to a_n referring to E_z defined in Eq. 4.25), \mathbf{h}_θ is the equivalent for H_θ (see eq. 4.25) and $\mathbf{G}(k)$ is a matrix

defining the relationship between the two.

In order to compute this relationship we must first calculate $E_z(r, \theta, z)$ and $H_\theta(r, \theta, z)$. From Eq. 4.20 & 4.10:

$$H_\theta(r, \theta, z) = \frac{-jkY_0}{\gamma_n} \sum_{n=0}^{\infty} a_n \cdot R'_n(\gamma_n r) \cdot \cos(i\theta) \cdot \cos(n\beta z) \quad (4.22)$$

where $\omega\epsilon = kY_0$

Which can be represented as :

$$E_z(r, \theta, z) = \sum_{n=0}^{\infty} e_n \cdot \cos(i\theta) \cdot \cos(n\beta z) \quad (4.23)$$

$$H_\theta(r, \theta, z) = \sum_{n=0}^{\infty} h_n \cdot \cos(i\theta) \cdot \cos(n\beta z) \quad (4.24)$$

where:

$$e_{zn} = a_n \cdot R_n(\gamma_n r) \quad (4.25)$$

$$h_{\theta n} = -a_n \cdot \frac{jkY_0}{\gamma_n} \cdot R'_n(\gamma_n r) \quad (4.26)$$

From this $\mathbf{G}(k)$ can be defined as a diagonal matrix:

$$h_{\theta n} = \mathbf{G}_{n,n}(k)e_{zn} = -jY_0 \frac{kR'_n(\gamma_n r)}{\gamma_n R_n(\gamma_n r)} e_n \quad (4.27)$$

as $\cos(i\theta) \cdot \cos(n\beta z)$ will cancel.

which can be arranged to give, at $r = r_1$:

$$\mathbf{G}_{n,n}(k) = -jY_0 \frac{kR'_n(\gamma_n r_1)}{\gamma_n R_n(\gamma_n r_1)} \quad (4.28)$$

now find $\mathbf{G}^I(k)$ (\mathbf{G} in region I) and $\mathbf{G}^{II}(k)$ (\mathbf{G} in region II) from Eq. 4.28, 4.18 and 4.19:

$$\mathbf{G}_{n,n}^I(k) = \frac{h_{\theta n}^I}{e_{zn}^I} = -jkY_0 \frac{\left(-J_{i+1}(\gamma_n r_1) + \frac{i}{r_1} J_i(\gamma_n r_1)\right)}{\gamma_n J_i(\gamma_n r_1)} \quad (4.29)$$

$$\mathbf{G}_{m,m}^{II}(k) = \frac{h_{\theta m}^{II}}{e_{zm}^{II}} \quad (4.30)$$

$$= -jkY_0 \frac{Y_i(\gamma_m r_2) \left(-J_{i+1}(\gamma_m r_1) + \frac{i}{\gamma_m r_1} J_i(\gamma_m r_1)\right) - J_i(\gamma_m r_2) \left(-Y_{i+1}(\gamma_m r_1) + \frac{i}{\gamma_m r_1} Y_i(\gamma_m r_1)\right)}{\gamma_m [Y_i(\gamma_m r_2) J_i(\gamma_m r_1) - J_i(\gamma_m r_2) Y_i(\gamma_m r_1)]} \quad (4.31)$$

where for instance e_{zn}^I is Eq. 4.25 for region I and m is the equivalent of n for region II .

4.2.4 Continuity Conditions (P and Q)

With infinite series the boundary conditions can be matched exactly for Eq. 4.23 and 4.24. In order to solve the equations the series must be finite, meaning that E_z and H_θ can no longer be matched at every point. Continuity of Poynting flux means that the stored energy in a loss-less cavity is constant which has been shown to be the correct matching condition.[66] Carter et. al. show that the continuity of the Poynting flux is satisfied using this method of field matching and series truncation.[66]

For TM modes:

$$\begin{array}{c} \text{Continuous at } r = r_1 \\ \overbrace{E_z \quad E_\theta \quad H_\theta} \end{array}$$

These matching conditions can be represented by two matrices $[P]$ and $[Q]$ which represent the E-field basis functions in one region in terms of the basis

functions in the other. These two equations are the same for modes with and without azimuthal variation as they just serve to represent the geometry rather than the fields themselves (although they are also derived from the boundary conditions, these apply equally in both regions).[66]

$$\mathbf{e}_z^{II} = \mathbf{P}\mathbf{e}_z^I \quad (4.32)$$

$$\mathbf{h}_\theta^I = \mathbf{Q}\mathbf{h}_\theta^{II} \quad (4.33)$$

where \mathbf{e}_z^I is a vector representation of 4.25 in region I , with \mathbf{e}_z^{II} in region II . \mathbf{h}_θ^I and \mathbf{h}_θ^{II} being the equivalent for H_θ (see eq. 4.24).

4.2.4.1 E-field Continuity P

For TM modes \mathbf{P} relates \mathbf{e}_z^I to \mathbf{e}_z^{II} , the field equations at the interface in this case are:

$$\begin{array}{ccc} \text{Region I} & & \text{Region II} \\ \sum_{n=0}^{\infty} e_{zn}^I \cos(n\beta^I z) dz & & \sum_{m=0}^{\infty} e_{zm}^{II} \cos(m\beta^{II} z) dz \quad 0 \leq z \leq z_1 \\ & & 0 \quad z_1 \leq z \leq z_2 \end{array}$$

as there can be no electric field in region I beyond $z = z_1$ (the height of the region) at $r = r_1$ and as there can be no transverse electric field on a metallic boundary E_z is also zero in region II . The azimuthal variation $\cos(i\theta)$ can safely be ignored as it is defined to be the same in both regions so matching is not required.

This leads to the continuity conditions at the interface ($r = r_1$):

$$\begin{array}{ll} E_z^I = E_z^{II} & 0 \leq z \leq z_1 \\ E_z^{II} = 0 & z_1 \leq z \leq z_2 \end{array}$$

This is integrated over the range over which continuity is required ($0 \leq z \leq z_1$), as for $z_1 \leq z \leq z_2$ E_z is zero. This yields the relationships:

$$\begin{aligned} \text{Region I} \quad \mathbf{e}_z^I \mathbf{P}^I &= \sum_{n=0}^N \int_0^{z_1} e_{zn}^I \cos(n\beta^I z) \cos(t\beta^{II} z) dz \\ \text{Region II} \quad \mathbf{e}_z^{II} \mathbf{P}^{II} &= \sum_{m=0}^M \int_0^{z_2} e_{zm}^{II} \cos(m\beta^{II} z) \cos(t\beta^{II} z) dz \end{aligned}$$

The factor of $\cos(t\beta^{II} z)$ (where t is an integer) forces the integral in region II to be zero except when n or $m = t$, yielding an invertible diagonal matrix.

This leads to:

$$\mathbf{e}_z^I \mathbf{P}^I = \mathbf{e}_z^{II} \mathbf{P}^{II} \quad (4.34)$$

$$\mathbf{e}_z^{II} = \mathbf{P}^I (\mathbf{P}^{II})^{-1} \mathbf{e}_z^I \quad (4.35)$$

$$\mathbf{e}_z^{II} = \mathbf{P} \mathbf{e}_z^I \quad (4.36)$$

where $\mathbf{P} = \mathbf{P}^I (\mathbf{P}^{II})^{-1}$.

4.2.4.2 H-field Continuity Q

Similarly for the H fields Q relates h_θ^I to h_θ^{II} , the field equations at the interface in this case are:

$$\begin{array}{ll} \text{Region I} & \text{Region II} \\ \sum_{n=0}^{\infty} h_{\theta n}^I \cos(n\beta^I z) dz & \sum_{m=0}^{\infty} h_{\theta m}^{II} \cos(m\beta^{II} z) dz \quad 0 \leq z \leq z_1 \\ & \text{Undefined} \quad z_1 \leq z \leq z_2 \end{array}$$

as the magnetic field in region I past $z = z_1$ (the height of the region) at $r = r_1$ is undefined by the boundary conditions. Again the azimuthal variation can be safely ignored.

The continuity conditions at the interface ($r = r_1$) are:

$$\begin{array}{ll} H_\theta^I = H_\theta^{II} & 0 \leq z \leq z_1 \\ H_\theta^{II} \text{ is undefined} & z_1 \leq z \leq z_2 \end{array}$$

This time the integration is performed over the range over which continuity is required ($0 \leq z \leq z_1$), as for $z_1 \leq z \leq z_2$ H_θ is undefined. The result is again

multiplied by $\cos(p\beta^I z)$ (where p is an integer) to obtain a diagonal matrix for region I . This gives the relationships:

$$\begin{aligned} \text{Region I} \quad \mathbf{h}_\theta^I \mathbf{Q}^I &= \begin{cases} \sum_{n=0}^N \int_0^{z_1} h_{\theta n}^I \cos(n\beta^I z) \cos(p\beta^I z) dz & \text{when } 0 \leq z \leq z_1 \\ 0 & \text{when } z > z_1 \end{cases} \\ \text{Region II} \quad \mathbf{h}_\theta^{II} \mathbf{Q}^{II} &= \sum_{m=0}^M \int_0^{z_1} h_{\theta m}^{II} \cos(m\beta^{II} z) \cos(p\beta z) dz \end{aligned}$$

Applying the matching conditions allows the definition of \mathbf{Q} :

$$\mathbf{h}_\theta^I \mathbf{Q}^I = \mathbf{h}^{II} \mathbf{Q}_\theta^{II} \quad (4.37)$$

$$\mathbf{h}_\theta^I = \mathbf{Q}^{II} \mathbf{Q}^{I-1} \mathbf{h}_\theta^{II} \quad (4.38)$$

$$\mathbf{h}_\theta^I = \mathbf{Q} \mathbf{h}^{II} \theta \quad (4.39)$$

4.2.5 Finding the Resonant Modes

The method used by Carter et al. [71] can be used to solve this set of equations. The four key matrix equations can now be set up as an eigensolution problem, which can be easily solved using a secant method or any zero finder. The solutions to this equation are k values which of course represent the resonant modes of the cavity.

$$\mathbf{h}_\theta^I = \mathbf{G}^I(k) \mathbf{e}_z^I \quad (4.40)$$

$$\mathbf{h}_\theta^{II} = \mathbf{G}^{II}(k) \mathbf{e}_z^{II} \quad (4.41)$$

$$\mathbf{h}_\theta^I = \mathbf{Q} \mathbf{h}_\theta^{II} \quad (4.42)$$

$$\mathbf{e}_z^{II} = \mathbf{P} \mathbf{e}_z^I \quad (4.43)$$

This can be rearranged in a similar solvable matrix equation:

$$\mathbf{e}_z^I = \mathbf{G}^I(k)^{-1} \mathbf{Q} \mathbf{G}^{II}(k) \mathbf{P}^{-1} \mathbf{e}_z^I \quad (4.44)$$

$$\left| \mathbf{G}^I(k)^{-1} \mathbf{Q} \mathbf{G}^{II}(k) \mathbf{P}^{-1} - \mathbf{I} \right| = 0 \quad (4.45)$$

4.2.6 Implementation

These equations are implemented in MathCAD 13.1 where solutions to Eq. 4.45 are found using the secant method which was terminated when the fractional error in was less than 10^{-12} . The full fields can be computed from this using the method described by Carter et. al. with the eigenvector found using the *eigenvec* MathCAD function equation on eq. 4.45 with an eigenvalue of 1. Similarly the other coefficients for the magnetic and electric fields can be found using Eq. 4.40, 4.41 & 4.43.[66] The stored energy Q and R/Q can be calculated from the resulting field distributions. Even in MathCAD (rather than an optimised FORTRAN implementation for instance) the calculations can take less than a second on a PC with a 3 GHz Pentium 4 processor and 4 GB of RAM (which is not required).

$$E_z(r, \theta, z) = \sum_{n=0}^N e_{zn} \cos(i\theta) \cos(n\beta z) \quad (4.46)$$

$$H_\theta(r, \theta, z) = \sum_{n=0}^M h_{\theta n} \cos(i\theta) \cos(n\beta z) \quad (4.47)$$

As mentioned, infinite series are not possible in a practical implementation, so the series must be truncated (see eq. 4.46). Carter et. al. show a useful relationship between N and M and the convergent behaviour of the system with respect to these variables.

$$M = \left\lceil \frac{z_2}{z_1} \cdot N \right\rceil \quad (4.48)$$

This does not necessarily translate entirely over to calculation of anything beyond the fundamental *TM* modes but the suggested relationships are implemented as a starting point.[66] This work does not investigate the effect on the convergent behaviour from the new terms.

A number of useful diagnostic techniques are used during this work to aid discovery of new modes. The secant method requires a starting value before it can begin to search for a root. As the independent variable is k the units for this starting value are m^{-1} . However this can be converted to H_z , where this frequency ideally being close to the expected resonant mode. Use of frequency rather than wave number is significantly more intuitive when searching for modes (see eq. 4.49).

$$f = \frac{kc}{2\pi} \quad (4.49)$$

Another technique is to plot Eq. 4.45 with respect to k or again more appropriately f . This allows the solutions to be seen graphically, reducing the possibility that solutions will be missed. This is especially useful as the secant method used in this situation is prone to overshooting solutions often by many GHz. The use of these plots allows the trial value to be set close to a solution, increasing the chances that the solver will converge as desired. The steps used for the secant method are much reduced compared to those necessary when only the fundamental mode is required. This is purely because the HOM are much more closely separated in the frequency space. Ideally the root finder would be able to find a full set of solutions in a given frequency range. This is not however a trivial task.

For certain calculations MathCAD is unable to represent numbers of a sufficient magnitude (over $1 \cdot 10^{307}$) which causes the simulation to fail. In the equations, such a large number is balanced by an equally large number on the bottom of the fraction. This problem could be solved by using large argument approximations of the Bessel functions when the arguments become too large,

but this is time consuming to derive and unnecessary at this stage, as the problem can often be avoided by altering the number of basis functions.

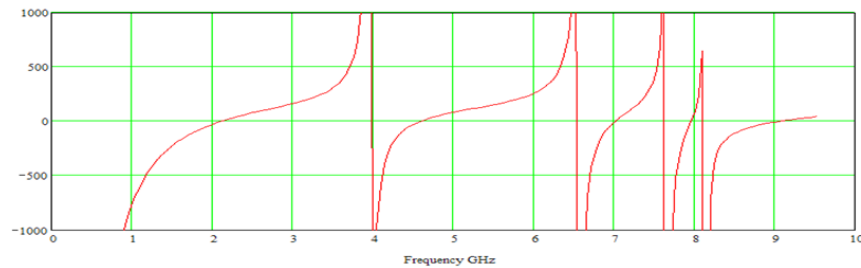


Figure 4.3: Eq. 4.45 displayed graphically, clearly showing solutions. These plots are for $i = 0$.

If these fields in both regions are plotted along the boundary ($r = r_1$), the matching of the fields can clearly be seen. The effect of truncating the series can be seen in the inexact matching of E_z between the two regions (see fig. 4.4a). Gibb's phenomenon can also be seen around $r = r_1$ in fig. 4.4a due to the sharp edge.

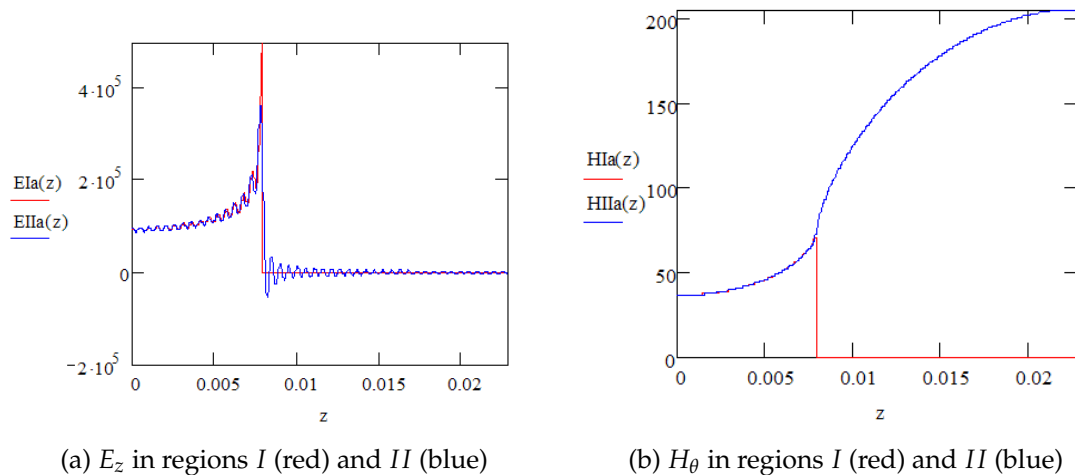


Figure 4.4: Plots of E_z and H_θ calculated by the Method of Moments along the interface between the 2 regions

4.3 *TM* Mode Results

The simulation of fundamental *TM* modes and their convergence in a re-entrant cavity has been investigated at length by Carter et al. [71]. However

there was no attempt made to use the same method to calculate modes of a higher frequency. The equations used to represent the fundamental TM are also able to represent all other modes with no azimuthal variation. Azimuthally varying modes are of course of particular interest with respect to MBK design. Simulating these HOM provides a starting point for the further extension of this method.

MAFIA 2D was chosen for the comparison as it is a well known and developmentally stable code. The 2D geometry has both speed and accuracy advantages over a 3D solver such as CST MWS, requiring fewer mesh cells and taking advantage of a further aspect of the geometry and field symmetry.

To calculate the HOM with no azimuthal variation i must be set to 0. This simplifies the expanded form of Eq. 4.40-4.43 to those used by Carter et al. [71] to calculate the fundamental modes.

Now the full solutions are known, modes with azimuthal variation can be found. This is of great interest for calculating the HOM of MBK cavities as many of the modes of interest have a non zero azimuthal variation index.

The cavity used in these tests has been used in previous testing of the method of moments and is a cavity it is known to model well.

Table 4.2: Test cavity dimensions (mm)

r_1	r_2	z_1	z_2
16	42	8	23

4.3.1 Azimuthal Variation Index (i) 0

The MAFIA results are calculated using a mesh of 200,000 mesh cells in the 2D solver specifically targeting TM modes. This constraint on the simulation greatly increases the accuracy and speed of the calculations. The errors are of the order of 0.01%, these results are calculated with the inclusion of the i terms.

Table 4.3: Comparison between frequencies (in GHz) from MAFIA and MoM Model for azimuthal variation (i) = 0

Mode	MoM	MAFIA	Err (MAFIA-MoM/MoM)
1	2.128	2.128	0.022%
2	4.623	4.623	0.011%
3	6.982	6.982	0.001%
4	7.989	7.989	0.001%
5	9.106	9.106	0.004%
6	11.038	11.038	0.002%
7	12.697	12.697	0.005%
8	13.241	13.240	0.004%
9	13.862	13.861	0.005%
10	14.491	14.491	0.003%

4.3.2 Azimuthal Variation Index (i) 1

When the azimuthal variation index (i) is increased to 1 the likelihood that the Bessel function arguments will overflow is increased. This makes it harder to find a number of basis functions which will allow the secant method to converge on a mode. The relationship between the number of basis functions in each region is only optimised for TM modes with no azimuthal variation, its performance where $i \neq 0$ is unknown.

The MAFIA 2D simulations no longer allow for the distinction between TE and TM modes as this distinction cannot be made. This results in many modes being calculated that are irrelevant and which have been omitted here.

Table 4.4: Comparison between frequencies (in GHz) from MAFIA and MoM Model for azimuthal variation (i) = 1

The modes highlighted in bold are potentially mixing with each other in pairs.

Mode	MoM	MAFIA	Err (MAFIA-MoM/MoM)
1	4.39	4.48	-2.05%
2	5.57	6.61	-15.76%
3	7.78	7.61	2.28%
4	8.47	8.33	1.66%
5	8.47	8.67	-2.32%
6	10.3	10.3	0.36%
7	10.3	11.3	-8.39%
8	12.0	12.1	-1.18%
9	13.1	13.2	-0.50%
10	13.7	13.5	0.93%

4.3.3 Inaccurate modes

A number of inaccurate modes are clearly present in the results so far. The first inaccurate mode is in fact a TE mode (see fig. 4.5) as shown in the example plot. This is not a mode which the code is designed to find; in fact it is specifically unable to calculate this mode accurately. One of the assumptions used when the G matrices are derived is that H_z is defined as 0 (which is of course true for TM modes) (see sec. 4.2).

The appearance of this mode in the results is not completely unexpected as MAFIA 2D, for instance, is also unable to differentiate between TE and TM modes when a azimuthal variation is specified.

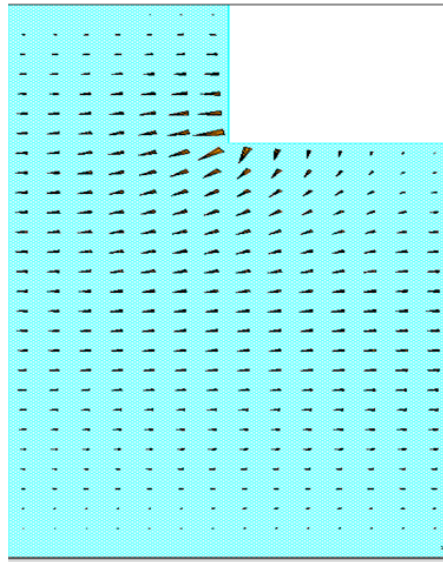


Figure 4.5: MAFIA plot of an inaccurate mode which is in fact a TE mode

4.4 TE Modes Derivation

As no trivial method to distinguish inaccurate modes has been found so far (short of a self defeating comparison with MAFIA 2D) the proposed solution is to allow the equations to fully represent all possible fields. The original method as proposed by Carter et al. [71] is only capable of calculating TM modes. However different starting assumptions can be used to allow the equations to

represent TE modes.

4.4.1 Field Component Equations

The TE modes must be calculated using a subtly different set of assumptions, the most important being that $H_z \neq 0$ and $E_z = 0$. This will lead to another similar set of matrix equations which can fully characterise TE modes, the starting point of which can be found from Ramo et al. [72] and is the equivalent of Eq. 4.2 (see eq. 4.50).

$$\nabla_t^2 H_z = -\gamma_n^2 H_z \quad (4.50)$$

with the solution:

$$H_z(r, \theta, z) = \left\{ \begin{array}{l} AJ_i(\gamma r) + BY_i(\gamma r) \\ AI_i(\gamma r) + BK_i(\gamma r) \end{array} \right\} \cdot \cos(i\theta) \quad (4.51)$$

where i = azimuthal variation index; A & B = constants; $\gamma = \sqrt{k^2 - k_z^2}$; $k_z = \omega/c$ in free space; k = wave number; H_z = magnetic field maximum in the z direction

To characterise the fields the field components must be found, which for a TE mode (remembering $E_z = 0$):

$$\overbrace{E_r \quad E_\theta \quad H_r \quad H_\theta \quad H_z}^{TE}$$

The components are from Eq. 4.5 to Eq. 4.8:

$$E_r = \frac{-j\omega\mu}{\gamma_n^2 r} \cdot \frac{\partial}{\partial \theta} H_z(r, \theta) \quad (4.52)$$

$$E_\theta = \frac{j\omega\mu}{\gamma_n^2} \cdot \frac{\partial}{\partial r} H_z(r, \theta) \quad (4.53)$$

$$H_r = \frac{j\beta}{\gamma_n^2} \cdot \frac{\partial}{\partial r} H_z(r, \theta) \quad (4.54)$$

$$H_\theta = \frac{j\beta}{\gamma_n^2 r} \cdot \frac{\partial}{\partial \theta} H_z(r, \theta) \quad (4.55)$$

4.4.2 Apply Boundary Conditions

Again these field components are only valid for a cylindrical waveguide. In the case of region *II* there are conducting boundaries at $z = 0$ & $z = z_2$. With the knowledge that the variation must again be sinusoidal, the z variation can be inferred from the boundary conditions.

Table 4.5: The new boundary conditions compared to the *TM* equivalents

	<i>TM</i>		<i>TE</i>	
	@ z_0 & z_2	z variation	@ z_0 & z_2	z variation
E_r	0	$\sin(n\beta z)$	0	$\sin(n\beta z)$
E_θ	0	$\sin(n\beta z)$	0	$\sin(n\beta z)$
E_z	$\neq 0$	$\cos(n\beta z)$	-	-
H_r	$\neq 0$	$\cos(n\beta z)$	$\neq 0$	$\cos(n\beta z)$
H_θ	$\neq 0$	$\cos(n\beta z)$	$\neq 0$	$\cos(n\beta z)$
H_z	-	-	0	$\sin(n\beta z)$
	$n = 0, 1 \dots \infty$		$n = 1, 2 \dots \infty$	
	$i = 0, 1 \dots \infty$		$i = 0, 1 \dots \infty$	

i now begins at 1 because there must always be a z variation in the H field as there must be no perpendicular magnetic field at the boundary. The full solution of the fields now becomes:

$$H_z(r, \theta, z) = R \cdot \cos(i\theta) \cdot \sin(n\beta z) \quad (4.56)$$

The derivation from this point closely follows the *TM* mode equation deri-

vation. With the exception of the boundary conditions which now are:

$$\frac{\partial H_z}{\partial r} = 0 @ r = 0 \quad (4.57)$$

$$\frac{\partial H_z}{\partial r} = 0 @ r = r_2 \quad (4.58)$$

4.4.3 Convert between E and H (\mathbf{G}^I and \mathbf{G}^{II})

\mathbf{G}^I and \mathbf{G}^{II} are derived as before. These are not reproduced here due to the increased complexity of the equations (see sec. B.1). The fields can still be expressed in terms of a summation:

$$H_z(r, \theta, z) = \sum_{n=0}^{\infty} h_{zn} \cdot \cos(i\theta) \cdot \sin(n\beta z) \quad (4.59)$$

$$E_\theta(r, \theta, z) = \sum_{n=0}^{\infty} e_{\theta n} \cdot \cos(i\theta) \cdot \sin(n\beta z) \quad (4.60)$$

where:

$$h_{zn} = a_n \cdot R_n(\gamma_n r) \quad (4.61)$$

$$e_{\theta n} = a_n \cdot \frac{j\omega\mu}{\gamma_n} \cdot R'_n(\gamma_n r) \quad (4.62)$$

4.4.4 Continuity Conditions (P & Q)

The different behaviours of the electric and magnetic fields at the surface of a conductor also require P and Q to be redefined.

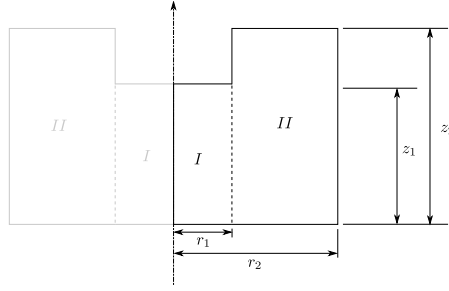


Figure 4.6: Recessed re-entrant cavity split into region *I* and region *II*

Continuity conditions at the interface compared to those of *TM* modes:

Continuous at $r = r_1$			
E_z	H_z	E_θ	H_θ
TM only	TE only	TE and TM	

4.4.4.1 H-field Continuity Q

Similarly to *TM* modes **Q** relates h^I to h^{II} but for H_z as opposed to H_θ . The field equations in this case are:

Region I	Region II	
$\sum_{n=0}^N h_{zn}^I \sin(n\beta^I z) dz$	$\sum_{m=0}^M h_{zm}^{II} \sin(m\beta^{II} z) dz$	$0 \leq z \leq z_1$
	Undefined	$z_1 \leq z \leq z_2$

The continuity conditions at the interface ($r = r_1$) are:

$$\begin{aligned}
 H_z^I &= H_z^{II} & 0 \leq z \leq z_1 \\
 H_z^{II} &\text{ is undefined} & z_1 \leq z \leq z_2
 \end{aligned}$$

Integrating over the range over which continuity is required ($0 \leq z \leq z_1$), as for $z_1 \leq z \leq z_2$ H_z is unknown. If this is multiplied by $\sin(t\beta^I z)$ (where t is an integer) the integral in region *I* is zero except when n or $m = t$, yielding a diagonal matrix.

$$\begin{aligned} \text{Region I} \quad \mathbf{h}^I \mathbf{Q}^I &= \sum_{n=0}^N h_{zn}^I \int_0^{z_1} \sin(n\beta^I z) \sin(t\beta^I z) dz \\ \text{Region II} \quad \mathbf{h}^{II} \mathbf{Q}^{II} &= \sum_{m=0}^M h_{zm}^{II} \int_0^{z_1} \sin(m\beta^{II} z) \sin(t\beta^I z) dz \end{aligned}$$

Where $[\mathbf{Q}^I]$ and $[\mathbf{Q}^{II}]$ are both t by n or m matrices with \mathbf{Q}^I being diagonal. The new \mathbf{Q} for TE modes can now be defined:

$$\mathbf{h}_z^I \mathbf{Q}^I = \mathbf{h}_z^{II} \mathbf{Q}^{II} \quad (4.63)$$

$$\mathbf{h}_z^I \mathbf{Q}^I \cdot (\mathbf{Q}^{II})^{-1} = \mathbf{h}_z^{II} \quad (4.64)$$

$$\mathbf{h}_z^{II} = \mathbf{Q} \cdot \mathbf{h}_z^I \quad (4.65)$$

4.4.4.2 E-field Continuity P

Similarly to TM modes \mathbf{P} relates e_{θ}^{II} to e_{θ}^I for E_{θ} (as opposed to E_z). H_{θ} is again continuous but for consistency in the matrix algebra it makes more sense to use E_{θ} for the continuity equations. The field equations in this case are:

$$\begin{array}{ccc} \text{Region I} & \text{Region II} & \\ \sum_{n=0}^N e_{\theta n}^I \sin(n\beta^I z) dz & \sum_{m=0}^M e_{\theta m}^{II} \sin(m\beta^{II} z) dz & 0 \leq z \leq z_1 \\ & 0 & z_1 \leq z \leq z_2 \end{array}$$

The continuity conditions at the interface ($r = r_1$) are again:

$$\begin{aligned} E_{\theta}^I &= E_{\theta}^{II} & 0 \leq z \leq z_1 \\ E_{\theta}^{II} &= 0 & z_1 \leq z \leq z_2 \end{aligned}$$

This can be integrated over the range where continuity is required ($0 \leq z \leq z_2$) as all fields are known. If this is multiplied by $\sin(p\beta z)$ (where p is an integer) the integral in region II is zero except when n or $m = p$, yielding a diagonal matrix.

$$\begin{aligned} \text{Region I} \quad \mathbf{e}_\theta^I [P^I] &= \sum_{n=0}^N e_{\theta n}^I \int_0^{z_1} \sin(n\beta^I z) \sin(p\beta^{II} z) dz \\ \text{Region II} \quad \mathbf{e}_\theta^{II} [P^{II}] &= \sum_{m=0}^M e_{\theta m}^{II} \int_0^{z_2} \sin(m\beta^{II} z) \sin(p\beta^{II} z) dz \end{aligned}$$

Where P^I and P^{II} are both t by n or m matrices with P^{II} being diagonal. The new P for TE modes can now be defined:

$$\mathbf{e}_\theta^{II} \mathbf{P}^{II} = \mathbf{e}_\theta^I \mathbf{P}^I \quad (4.66)$$

$$\mathbf{e}_\theta^{II} \mathbf{P}^{II} (\mathbf{P}^I)^{-1} = \mathbf{e}_\theta^I \quad (4.67)$$

$$\mathbf{e}_\theta^I = \mathbf{P} \cdot \mathbf{e}_\theta^{II} \quad (4.68)$$

4.4.5 Finding the Resonant Modes

As the full set of matrices for TE modes have been found:

$$\mathbf{h}_z^I = \mathbf{G}^I(k) \cdot \mathbf{e}_\theta^I \quad (4.69)$$

$$\mathbf{h}_z^{II} = \mathbf{G}^{II}(k) \cdot \mathbf{e}_\theta^{II} \quad (4.70)$$

$$\mathbf{h}_z^{II} = \mathbf{Q} \cdot \mathbf{h}_z^I \quad (4.71)$$

$$\mathbf{e}_\theta^I = \mathbf{P} \cdot \mathbf{e}_\theta^{II} \quad (4.72)$$

again a solvable matrix equation can be set up:

$$\mathbf{h}_z^I = \mathbf{G}^I(k) \cdot \mathbf{P} \cdot \mathbf{G}^{II}(k)^{-1} \cdot \mathbf{Q} \cdot \mathbf{h}_z^I \quad (4.73)$$

$$\left| \mathbf{G}^I(k) \cdot \mathbf{P} \cdot \mathbf{G}^{II}(k)^{-1} \cdot \mathbf{Q} - \mathbf{I} \right| = 0 \quad (4.74)$$

4.5 TE Mode Results

The results obtained compare very well with MAFIA. The table below combines the results from the TM and the TE basis function simulations. The modes found are calculated with a good accuracy all mostly below 0.5% (see table. 4.7). The large argument approximations have not been calculated so the number of basis functions (and ultimately the accuracy) is limited.

Table 4.6: Test cavity dimensions (mm)

r_1	r_2	z_1	z_2
5	10	15	20

4.5.1 Azimuthal Variation Index (i) 1

Table 4.7: Full table of results from TE and TM basis function simulations for azimuthal index (m) 1

Mode	MAFIA ⁵ f (GHz)	MoM f TE (GHz)	MoM f TM (GHz)	Err (MAFIA-MoM/MoM) f (GHz)
1	11.395			
2	15.193			
3	18.766		18.688	0.42%
4	21.315			
5	23.242			
6	27.004	27.167		-0.60%
7	27.597		27.636	-0.14%
8	30.704	30.913		-0.68%
9	31.582	31.797		-0.68%
10	32.028		32.027	0.00%
11	33.763		33.769	-0.02%
12	34.317		34.428	-0.32%
13	35.911		36.018	-0.30%
14	36.244	36.778		-1.45%
15	39.174			
16	39.538		39.515	0.06%
17	40.217	40.048		0.42%
18	41.935	41.950		-0.04%
19	42.710			
20	44.500		44.488	0.03%

The results no longer contain R/Q and Q information as the equations used to calculate these are closely related to the equations used to plot the full field

distributions.

4.6 Coupled Modes Derivation

Sections 4.3 & 4.5 show that defining modes purely as TE or TM is not adequate for a full solution. In a re-entrant cavity all the modes are a linear combination of TE and TM modes, some modes have a higher amplitude of TE or TM modes and these can be modelled successfully as either TE or TM . Other modes are neither pure TE or pure TM and must be modeled as a mixture of both.

As TE and TM only include either H_z and E_z respectively the \mathbf{G} matrices previously derived are still valid. However as E_θ and H_θ are shared between the modes the matching equations are no longer valid. These are the key matrices for mixing the two types of mode. As both TE and TM are required the full set of boundary conditions are needed:

Table 4.8: The boundary conditions of TM and TE modes (both are required)

	TM		TE	
	@ z_0 & z_2	z variation	@ z_0 & z_2	z variation
E_r	0	$\sin(n\beta z)$	0	$\sin(n\beta z)$
E_θ	0	$\sin(n\beta z)$	0	$\sin(n\beta z)$
E_z	$\neq 0$	$\cos(n\beta z)$	-	-
H_r	$\neq 0$	$\cos(n\beta z)$	$\neq 0$	$\cos(n\beta z)$
H_θ	$\neq 0$	$\cos(n\beta z)$	$\neq 0$	$\cos(n\beta z)$
H_z	-	-	0	$\sin(n\beta z)$
	$n = 0, 1 \dots \infty$		$n = 1, 2 \dots \infty$	
	$i = 0, 1 \dots \infty$		$i = 0, 1 \dots \infty$	

4.6.1 Mixing E_θ

The matching equations for E_θ , instead of being based on a single set of basis functions, both TE and TM are required. The fields are now represented as the summation of $e_z^{x(TE)}$ & $e_z^{x(TM)}$:

$$\begin{array}{ccc}
\text{Region I} & & \text{Region II} \\
\sum_n^N \left(e_{\theta n}^{I(TM)} + e_{\theta n}^{I(TE)} \right) \sin(n\beta^I z) & & \sum_m^M \left(e_{\theta m}^{II(TM)} + e_{\theta m}^{II(TE)} \right) \sin(m\beta^{II} z) \quad 0 \leq z \leq z_1 \\
& & 0 \quad z_1 \leq z \leq z_2
\end{array}$$

The continuity conditions at the interface ($r = r_1$) do not change and remain:

$$\begin{aligned}
E_{\theta}^I &= E_{\theta}^{II} & 0 \leq z \leq z_1 \\
E_{\theta}^{II} &= 0 & z_1 \leq z \leq z_2
\end{aligned}$$

These equations must again be integrated over the full range of z as E_{θ} remains fully defined (see sec. 4.2.4.1). To diagonalise the matrix both sides are multiplied by $\sin(p\beta^{II}z)$ to make region *I* diagonal (and invertible).

$$\begin{aligned}
\text{Region I} \quad & \left(\mathbf{e}_{\theta}^{I(TM)} + \mathbf{e}_{\theta}^{I(TE)} \right) \mathbf{A}^I = \\
& \sum_n^N \left(e_{\theta n}^{I(TM)} + e_{\theta n}^{I(TE)} \right) \int_0^{z_1} \sin(n\beta^I z) \sin(p\beta^{II} z) dz \\
\text{Region II} \quad & \left(\mathbf{e}_{\theta}^{II(TM)} + \mathbf{e}_{\theta}^{II(TE)} \right) \mathbf{A}^{II} = \\
& \sum_m^M \left(e_{\theta m}^{II(TM)} + e_{\theta m}^{II(TE)} \right) \int_0^{z_2} \sin(m\beta^{II} z) \sin(p\beta^{II} z) dz
\end{aligned}$$

The continuity conditions allow us to equate both region *I* and *II*.

$$\mathbf{A}^I \left(\mathbf{e}_{\theta}^{I(TM)} + \mathbf{e}_{\theta}^{I(TE)} \right) = \mathbf{A}^{II} \left(\mathbf{e}_{\theta}^{II(TM)} + \mathbf{e}_{\theta}^{II(TE)} \right) \quad (4.75)$$

With \mathbf{A}^I and \mathbf{A}^{II} defined as follows:

$$\mathbf{A}^I = \int_0^{z_1} \sin(n\beta^I z) \sin(p\beta^{II} z) dz \quad (4.76)$$

$$\mathbf{A}^{II} = \int_0^{z_2} \sin(m\beta^{II} z) \sin(p\beta^{II} z) dz \quad (4.77)$$

Usefully this is equivalent to \mathbf{P}^{TE} which will be used in its place from this point (see sec. 4.4.4.2):

$$\mathbf{A} = \frac{\mathbf{A}^I}{\mathbf{A}^{II}} = \mathbf{P}^{TE} \quad (4.78)$$

4.6.2 Mixing H_θ

H_θ can be matched in the same way using the sum of both TE and TM basis functions:

$$\begin{array}{ccc} \text{Region I} & & \text{Region II} \\ \sum_n^N \left(h_{\theta n}^{I(TM)} + h_{\theta n}^{I(TE)} \right) \cos(n\beta^I z) & \sum_m^M \left(h_{\theta m}^{II(TM)} + h_{\theta m}^{II(TE)} \right) \cos(m\beta^{II} z) & 0 \leq z \leq z_1 \\ & 0 & z_1 \leq z \leq z_2 \end{array}$$

The continuity conditions at the interface ($r = r_1$) are again:

$$\begin{array}{ll} H_\theta^I = H_\theta^{II} & 0 \leq z \leq z_1 \\ H_\theta^{II} \text{ undefined} & z_1 \leq z \leq z_2 \end{array}$$

As the fields are only defined between 0 & z_1 this is used as the integration range. Again one side must be diagonalised (and so made invertible) by multiplying by $\cos(p\beta^I z)$.

$$\begin{array}{l} \text{Region I} \quad \left(\mathbf{h}_\theta^{I(TM)} + \mathbf{h}_\theta^{I(TE)} \right) \mathbf{B}^I = \sum_n^N \left(h_{\theta n}^{I(TM)} + h_{\theta n}^{I(TE)} \right) \int_0^{z_1} \cos(n\beta^I z) \cos(p\beta^I z) dz \\ \text{Region II} \quad \left(\mathbf{h}_\theta^{II(TM)} + \mathbf{h}_\theta^{II(TE)} \right) \mathbf{B}^{II} = \sum_m^M \left(h_{\theta m}^{II(TM)} + h_{\theta m}^{II(TE)} \right) \int \cos(m\beta^{II} z) \cos(p\beta^I z) dz \end{array}$$

Representing the integrals as matrices and performing the matching yields:

$$\mathbf{B}^I \left(\mathbf{h}_\theta^{I(TM)} + \mathbf{h}_\theta^{I(TE)} \right) = \mathbf{B}^{II} \left(\mathbf{h}_\theta^{II(TM)} + \mathbf{h}_\theta^{II(TE)} \right) \quad (4.79)$$

With B^I & B^{II} defined as:

$$\mathbf{B}^I = \int_0^{z_1} \cos(n\beta^I z) \cos(p\beta^I z) \quad (4.80)$$

$$\mathbf{B}^{II} = \int_0^{z_1} \cos(m\beta^{II} z) \cos(p\beta^I z) \quad (4.81)$$

Usefully this is the equivalent of Q^{TM} (see sec. 4.2.4.2) which is used in its place from this point:

$$\mathbf{A} = \frac{\mathbf{B}^{II}}{\mathbf{B}^I} = \mathbf{Q}^{TM} \quad (4.82)$$

4.6.3 Convert Between θ and z (H & J)

For this formulation two more matrices are needed, H & J , to convert between the θ and z components in the same region. They are derived using the same method as the G matrices and are defined as follows:

$$\mathbf{e}_\theta^{I(TM)} = \frac{E_\theta}{E_z} \mathbf{e}_z^{I(TM)} \quad (4.83)$$

$$\mathbf{e}_\theta^{I(TM)} = \left[H^I \right] \mathbf{e}_z^{I(TM)} \quad (4.84)$$

For the magnetic field:

$$\mathbf{h}_\theta^{I(TE)} = \frac{H_\theta}{H_z} \mathbf{h}_z^{I(TE)} \quad (4.85)$$

$$\mathbf{h}_\theta^{I(TE)} = \left[J^I \right] \mathbf{h}_z^{I(TE)} \quad (4.86)$$

4.6.4 Recap

There are 12 variables $\mathbf{e}_z^{x(TM)}$, $\mathbf{h}_z^{x(TE)}$, $\mathbf{e}_\theta^{x(TM)}$, $\mathbf{e}_\theta^{x(TE)}$, $\mathbf{h}_\theta^{x(TM)}$, $\mathbf{h}_\theta^{x(TE)}$ in both regions, as a result 12 simultaneous equations are needed.

The basis of this solution are the two mixing equations from Eq. 4.80-4.82 and Eq. 4.76-4.78:

$$\mathbf{P}^{TE} \left(\mathbf{e}_\theta^{I(TM)} + \mathbf{e}_\theta^{I(TE)} \right) = \left(\mathbf{e}_\theta^{II(TM)} + \mathbf{e}_\theta^{II(TE)} \right) \quad (4.87)$$

$$\left(\mathbf{h}_\theta^{I(TM)} + \mathbf{h}_\theta^{I(TE)} \right) = \mathbf{Q}^{TM} \left(\mathbf{h}_\theta^{II(TM)} + \mathbf{h}_\theta^{II(TE)} \right) \quad (4.88)$$

There are also the matching equations for E_z and H_z from Sec. 4.2.4.1 & 4.4.4.1:

$$\mathbf{e}_z^{II(TM)} = \mathbf{P}^{TM} \cdot \mathbf{e}_z^{I(TM)} \quad (4.89)$$

$$\mathbf{h}_z^{I(TE)} = \mathbf{Q}^{TE} \cdot \mathbf{h}_z^{II(TE)} \quad (4.90)$$

The G matrices to convert between E_z, H_z and their associated θ components.

$$\mathbf{h}_\theta^{I(TM)} = \mathbf{G}^{ITM} \cdot \mathbf{e}_z^{I(TM)} \quad (4.91)$$

$$\mathbf{e}_\theta^{I(TE)} = \mathbf{G}^{ITE} \cdot \mathbf{h}_z^{I(TE)} \quad (4.92)$$

Additionally the H & J matrices convert between the θ and z components in the same region. They are derived using the same method as the G matrices.

$$\mathbf{e}_\theta^{I(TM)} = \mathbf{H}^I \cdot \mathbf{e}_z^{I(TM)} \quad (4.93)$$

$$\mathbf{h}_\theta^{I(TE)} = \mathbf{J}^I \cdot \mathbf{h}_z^{I(TE)} \quad (4.94)$$

4.6.5 Solution

The basis of this solution are the mixing equations for H_θ and E_θ :

$$\mathbf{P}^{TE} \left(\mathbf{e}_\theta^{I(TM)} + \mathbf{e}_\theta^{I(TE)} \right) = \left(\mathbf{e}_\theta^{II(TM)} + \mathbf{e}_\theta^{II(TE)} \right) \quad (4.95)$$

$$\left(\mathbf{h}_\theta^{I(TM)} + \mathbf{h}_\theta^{I(TE)} \right) = \mathbf{Q}^{TM} \left(\mathbf{h}_\theta^{II(TM)} + \mathbf{h}_\theta^{II(TE)} \right) \quad (4.96)$$

Into which H^x and G^x can be substituted in to convert from e_θ and h_θ to e_z and h_z :

$$\mathbf{P}^{TE} \left(\mathbf{H}^I \mathbf{e}_z^{I(TM)} + \mathbf{G}^{I(TE)} \mathbf{h}_z^{I(TE)} \right) = \left(\mathbf{H}^{II} \mathbf{e}_z^{II(TM)} + \mathbf{G}^{II(TM)} \mathbf{h}_z^{II(TE)} \right) \quad (4.97)$$

$$\left(\mathbf{G}^{I(TM)} \mathbf{e}_z^{I(TM)} + \mathbf{J}^I \mathbf{h}_z^{I(TE)} \right) = \mathbf{Q}^{TM} \left(\mathbf{G}^{II(TM)} \mathbf{e}_z^{II(TM)} + \mathbf{J}^{II} \mathbf{h}_z^{II(TE)} \right) \quad (4.98)$$

These equations can be converted to matrix form:

$$\begin{bmatrix} \mathbf{e}_z^{I(TM)} \\ \dots \\ \mathbf{h}_z^{I(TE)} \end{bmatrix} \begin{bmatrix} \mathbf{P}^{TE} \cdot \mathbf{H}^I & \vdots & \mathbf{P}^{TE} \cdot \mathbf{G}^{I(TE)} \\ \dots & & \dots \\ \mathbf{G}^{I(TM)} & \vdots & \mathbf{J}^I \end{bmatrix} = \quad (4.99)$$

$$\begin{bmatrix} \mathbf{H}^{II} & \vdots & \mathbf{G}^{II(TM)} \\ \dots & & \dots \\ \mathbf{Q}^{TM} \cdot \mathbf{G}^{II(TM)} & \vdots & \mathbf{Q}^{TM} \cdot \mathbf{J}^{II} \end{bmatrix} \begin{bmatrix} \mathbf{e}_z^{II(TM)} \\ \dots \\ \mathbf{h}_z^{II(TE)} \end{bmatrix}$$

Eq 4.89 & 4.90 are still missing, which can be included as follows to:

$$\begin{bmatrix} \mathbf{e}_z^{I(TM)} \\ \dots \\ \mathbf{h}_z^{II(TE)} \end{bmatrix} \begin{bmatrix} \mathbf{P}^{TE} \cdot \mathbf{H}^I & \vdots & \mathbf{P}^{TE} \cdot \mathbf{G}^{I(TE)} \cdot \mathbf{Q}^{TE} \\ \dots & & \dots \\ \mathbf{G}^{I(TM)} & \vdots & \mathbf{J}^I \cdot \mathbf{Q}^{TE} \end{bmatrix} = \quad (4.100)$$

$$\begin{bmatrix} \mathbf{H}^{II} \cdot \mathbf{P}^{TM} & \vdots & \mathbf{G}^{II(TE)} \\ \dots & & \dots \\ \mathbf{Q}^{TM} \cdot \mathbf{G}^{II(TM)} \cdot \mathbf{P}^{TM} & \vdots & \mathbf{Q}^{TM} \cdot \mathbf{J}^{II} \end{bmatrix} \begin{bmatrix} \mathbf{e}_z^{I(TM)} \\ \dots \\ \mathbf{h}_z^{II(TE)} \end{bmatrix}$$

In order to solve this the equation can be divided through by the partition vector:

$$\begin{aligned}
 & \left[\begin{array}{cc} \mathbf{P}^{TE} \cdot \mathbf{H}^I & \vdots \quad \mathbf{P}^{TE} \cdot \mathbf{G}^{I(TE)} \cdot \mathbf{Q}^{TE} \\ \dots & \dots \\ \mathbf{G}^{I(TM)} & \vdots \quad \mathbf{J}^I \cdot \mathbf{Q}^{TE} \end{array} \right] = \\
 & \left[\begin{array}{cc} \mathbf{H}^{II} \cdot \mathbf{P}^{TM} & \vdots \quad \mathbf{G}^{II(TE)} \\ \dots & \dots \\ \mathbf{Q}^{TM} \cdot \mathbf{G}^{II(TM)} \cdot \mathbf{P}^{TM} & \vdots \quad \mathbf{Q}^{TM} \cdot \mathbf{J}^{II} \end{array} \right] = 0
 \end{aligned} \tag{4.101}$$

This can be solved for k (an argument of the \mathbf{G} , \mathbf{H} and \mathbf{J} matrices) in the usual way.

4.7 Mixed Mode Results

The results presented here are those which match modes found using MA-FIA. Accompanying these are a number of unphysical modes, these are identified and discarded by comparison with MAFIA.

4.7.1 Azimuthal Variation Index (i) 0

The calculation of TM modes remains accurate with the new formulation, and all modes are calculated to a high degree of precision.

Table 4.9: Comparison between frequencies (in GHz) for Azimuthal Variation (i) = 0 from MAFIA and MoM Model with pure TM errors from table 4.3

Mode	MoM	MAFIA	Err (MAFIA-MoM/MoM)	
			Mixed	TM (see table. 4.3)
1	2.128396	2.128	0.020%	0.022%
2	4.623091	4.623	0.009%	0.011%
3	6.982075	6.982	0.001%	0.001%
4	7.989248	7.989	0.000%	0.001%
5	9.106587	9.106	0.006%	0.004%
6	11.038251	11.038	0.003%	0.002%
7	12.697287	12.697	0.005%	0.005%
8	13.24067	13.240	0.003%	0.004%
9	13.86170	13.861	0.004%	0.005%
10	14.491119	14.491	0.002%	0.003%

4.7.2 Azimuthal Variation Index (i) 1

The use of the mixed equations act to increase the accuracy of the $i = 1$ modes. For instance modes 1, 2, 3 and 7 have increased in accuracy by a significant margin. Mode 2 has moved closer to the correct value with an error of 4.81% from 15.76%, although this is still unacceptably large. The method still has difficulty separating modes which are close together in frequency. A mode is found between modes 4 and 5 as was the case for the pure TM basis functions.

In the pure TM formulation only one mode was found with the MoM for modes 6 and 7 found by MAFIA. The mixed mode formulation discovers these modes separately with an accuracy better than 1%. Although the accuracy of mode 6 decreases from 0.36% to 0.79%.

The spurious modes are, at this stage, unexplained. They could be related to inaccuracies due to truncation or perhaps the method of solution of the matrix equations, further work is required to identify their source.

Table 4.10: Comparison between frequencies (in GHz) for Azimuthal Variation (i) = 1 from MAFIA and MoM Model with pure TM errors from table 4.4

Mode	MoM	MAFIA	Err (MAFIA-MoM/MoM)	
			Mixed	TM (see table. 4.4)
1	4.492241	4.48	0.26%	-2.05%
2	6.307777	6.61	-4.81	-15.76%
3	7.628712	7.61	0.25%	2.28%
4	8.442429	8.33	1.30%	1.66%
5	8.442429	8.67	-2.72%	-2.32%
6	10.373085	10.3	0.79%	0.36%
7	11.225034	11.3	0.43%	-8.39%
8	11.96609	12.1	1.32%	-1.18%
9	13.204762	13.2	0.08%	-0.50%
10	13.597467	13.5	0.42%	0.93%

4.8 Conclusions

A HOM calculation solution based on the work by Carter et al. [71] on the method of moments in order to avoid the problems found in using FE or FDTD solvers shows promise.[66] With no alteration the equations which represent fundamental TM modes also calculate modes with no azimuthal variation to a comparable accuracy.

The re-derivation of the TM functions including the $\cos(i\theta)$ term allows the TM modes to be calculated with an accuracy of 0.01. However for modes with $i \neq 0$ the distinction between TE and TM becomes blurred. In fact there are no longer TE and TM modes but only hybrid modes some of which look nearly like TE , some like TM and some like neither. As a result of this modes which are in fact TE modes appear in the solutions. As a fundamental assumption of these equations $H_z = 0$, this is not true for TE modes and as a result there modes are imperfectly calculated. This adds a number of highly erroneous modes to the otherwise well calculated TM modes.

The solution to this is to augment the TM calculations with those for the TE mode. Using the same derivation and formulation of equations these can also be calculated to a high degree of accuracy. Combining the two sets of equations

yields a solution capable of calculating many, but not all modes. Modes which are not clearly TM or TE are not represented in this set.

What is needed is a combination of both these types of mode which can fully represent modes with non zero H_z and E_z . This is achieved by altering the equations that match the fields at the interface. As these equations match E_θ and H_θ , which appear in both TM and TE they can be used to mix the two types of mode. This means that a linear combination of TM and TE can meet the boundary and matching conditions without either set fully matching either.

Mixed mode calculations maintain accuracy for TM modes with 0 azimuthal variation and increases the accuracy for an azimuthal variation of 1. The accuracy of some of these modes is however too low to be of use at almost 5%. These good results are interspersed with a number of spurious modes. Further work is required to assess the issues with this method and propose solutions.

If only modes with $i = 0$ are required, the method of moments is a rapid and accurate method for determining cavity properties. The solution for modes with $i \neq 0$ is more complex and can yield accurate results but they are combined with poor quality solutions. In a single beam klystron, as only modes with an appreciable field along the z axis (and so high R/Q) are of interest, these TE -like modes could be filtered out. Unfortunately an MBK has off axis beams so this is not an available solution and the full mixed TE and TM formulation must be used.

Chapter 5

Wake-field Stability Analysis

5.1 Introduction

Traditional klystron stability analysis relates purely to a single beam. Phenomena such as monotron oscillations are investigated only in their effects on a single beam [2, 51]. Although of course if monotron oscillations are avoided for a single beam they will also be avoided in the multiple case.

MBKs are sometimes stated to be more stable [60] than their single beam cousins but the literature contains little analysis of effects due to beam current spread or beam loss. The chances of a beam being lost are at least equal to those of a single beam klystron multiplied by the number of beams. In reality the relative complexity of the wiring and heater structures combined with a shared vacuum vessel may increase the probability beyond this. For instance heater wiring failure was believed to be the cause of at least one lost MBK.[74]

In single beam klystrons the beam is situated in the centre of the cavity. This is neatly at the position of at least one of the null points of the majority of modes. The dipole mode for instance will not interact with a central beam, in fact any mode with a theta variation will only couple very weakly. The only modes of interest are the higher harmonics of the fundamental mode.

In a multiple beam klystron the situation is much more complex. As the multiple beams by their very nature are off centre almost all modes in the cavity

can be excited. The dipole mode in the coaxial cavity (see sec. 2.11.2) has been shown to be particularly close in frequency to the operating mode with a high maximum R/Q . This R/Q is of course not constant for all of the beams so a simple impression of the effect of this in a practical tube is not clear .

In this section it is assumed that the modulation of all the beams is perfectly in phase, the result of a well designed input and output cavity. The perfect MBK has beams which are all modulated in phase with one another and with equal current. With these assumptions the dipole mode cannot be excited as, when there are an even number of beams, the induced voltages will cancel. The interaction on one side of the mode will be perfectly canceled out by the other half.

Unfortunately all cathodes are not created equal and individual cathodes have different relationships between heater power, voltage and beam current which differ from their design targets due to the complexity of the surface chemistry. Adjustments to the beam current can be made using the heater power up to a point, but only if the cathode heaters are electrically independent.[34] Cathodes also do not degrade equally; a perfectly matched cathode at production may be poorly matched after use.

Individual heater currents are difficult to achieve as the heaters are often wired in parallel although the technical barriers associated with this are relatively insignificant (each heater supply would need a high voltage isolating transformer).[74] Instead the cathodes are carefully selected at the factory to be well matched. If the spread of currents differs from the perfect MBK's assumption of 0% spread (as it does) the dipole (and other theta varying modes) are excited.

Altering the individual gun voltages is problematic due to voltage generation issues (each gun would require either its own high tension power supply or some kind of high voltage step down circuit). Quite apart from this difficulty the beam velocities would differ, causing phase problems further down

the interaction structure.

5.2 Time domain

5.2.1 Theory

This chapter attempts to model an output cavity of a multiple beam klystron using standard wake-field theory to model it as a series of independent gaps with a geometrical relationship. It makes no attempt to address any effect of gain through the structure, transient effects or space-charge based cross talk between the beams.

These limitations aside, investigation of the beam current spread is possible, along with the effect of mode characteristics.

The operating frequency and dipole frequencies are respectively:

$$f_1 = \frac{\omega_1}{2\pi} \quad (5.1)$$

$$f_2 = \frac{\omega_2}{2\pi} \quad (5.2)$$

from which the assumed temporal bunch length and bunch separation for the beam can be found:

$$\sigma_b = \frac{\pi}{3\omega} \quad (5.3)$$

$$\tau_b = \frac{2\pi}{\omega} \quad (5.4)$$

with the bunch in the form of a Gaussian $\exp\left(-\frac{t^2}{\sigma_b^2}\right)$ and the damping factor of each mode [75]:

$$\tau_d(i) = \left(\frac{2Q_i}{\omega_i}\right) \quad (5.5)$$

5.2.1.1 Beam Current (I_b) Variation

When the beam currents are the same and in perfect phase all of the dipole components cancel and the output is pure monopole. This is of course non physical as perfect matching of the guns is impossible to achieve. Thales report that a spread of currents (Δ_b) of as much as 5% of I_b is expected between the beams and unless otherwise stated this value shall be used. A normal distribution is used to represent this current variation with the target percentage variation set as the 3rd standard deviation ($\sigma = \Delta_b/3$) (see fig. 5.1).

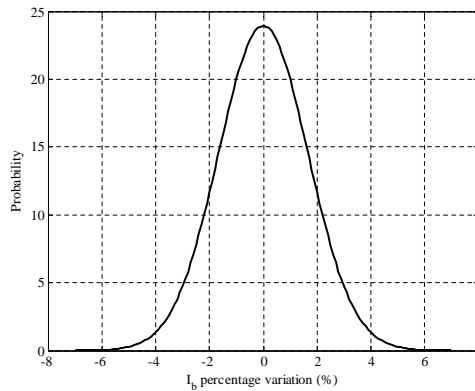


Figure 5.1: Normal probability distribution of percentage variation of I_b between the beams

In order to take into account the variation of the electron guns the beam current of the i^{th} beam is defined to be:

$$I_b(i) = 1A + \delta_i \cdot 1A \quad (5.6)$$

where δ_i = a Gaussian random number with a standard deviation of $\sigma = \Delta_b/3$ ($\Delta_b = 5\% \cdot I_b$ in this case) and is unique for each beam.

5.2.1.2 Wake-fields

We assume for simplicity that the bunch has a Gaussian current distribution and the spectral content F_i is given by:

$$F_i = \int i(t)e^{-j\omega t} dt = \exp\left(-\frac{1}{2}\left(\frac{\omega_i\sigma_b}{2\pi}\right)^2\right) \quad (5.7)$$

The real¹ parallel field generated in the wake of the bunch for each mode by a single bunch can be found [75] and from this the change in energy stored in the cavity qW_{\parallel} can be calculated, where W_{\parallel} is the wake potential, the average Lorentz force on a charge q travelling through a gap.

$$W_{\parallel}(i) = \omega_i \cdot F_i \cdot \left(\frac{R}{Q}\right)_i \cdot \cos(\omega_i\tau_b) \cdot \exp\left(\frac{-\tau_b}{\tau_d(i)}\right) \quad (5.8)$$

where $\frac{R}{Q}_i$ is the $\frac{R}{Q}$ of the i^{th} mode and ω_i is the angular frequency i^{th} mode

For a train of bunches (as found in a klystron) the wakes to infinity (steady state) must be summed:

$$W_{\parallel}(i) = \omega_i \cdot F_i \cdot \left(\frac{R}{Q}\right)_i \cdot \sum_{n=1}^{\infty} \cos(\omega_i\tau_b n) \exp\left(\frac{-\tau_b n}{\tau_d(i)}\right) \quad (5.9)$$

where $\tau_b n$ is the temporal position of each bunch.

The sum to infinity of eq. 5.9 can be found using the identity [76, 77]:

$$\sum_{n=1}^{\infty} \cos(nw) \cdot \exp(-nd) \equiv \frac{\sinh(d)}{2 \cosh(d) - 2 \cos(w)} - \frac{1}{2} \quad (5.10)$$

which applied to eq. 5.9 yields:

$$W_{\parallel}(i) = \omega_i \cdot F_i \cdot \left(\frac{R}{Q}\right)_i \cdot \left[\frac{\sinh\left(\frac{\tau_b}{\tau_d(i)}\right)}{2 \cosh\left(\frac{\tau_b}{\tau_d(i)}\right) - 2 \cos(\omega_i\tau_b)} - \frac{1}{2} \right] \quad (5.11)$$

As each bunch only “sees” half the voltage it induces, half the energy contributed by the zeroth bunch is added to eq. 5.11:

¹it is assumed the imaginary component is small for strong beam interactions as the real part is the energy lost by the beam

$$\omega_i \cdot F_i \cdot \left(\frac{R}{Q}\right)_i \cdot \frac{1}{2} \quad (5.12)$$

giving:

$$W_{\parallel}(i) = \omega_i \cdot F_i \cdot \left(\frac{R}{Q}\right)_i \cdot \frac{\sinh\left(\frac{\tau_b}{\tau_d(i)}\right)}{2 \cosh\left(\frac{\tau_b}{\tau_d(i)}\right) - 2 \cos(\omega_i \tau_b)} \quad (5.13)$$

As the dipole mode has two positions the x and y directions are arbitrarily defined and the interaction between each beam and each orientation are found. The dipole components excited by each beam can be found using:

$$X = \sum_{i=1}^{N_b} \left(W_{\parallel}(2) \cdot I_b(n) \cdot \cos(\theta_n) \right) \quad (5.14)$$

$$Y = \sum_{i=1}^{N_b} \left(W_{\parallel}(2) \cdot I_b(n) \cdot \sin(\theta_n) \right) \quad (5.15)$$

where $I_b(n)$ is the current and θ_i the angle of the n^{th} beam and $W_{\parallel}(2)$ is the wake-field for the dipole mode.

And the monopole component from:

$$M = \sum_{i=1}^{\infty} \left(W_{\parallel}(1) \cdot I_b(n) \right) \quad (5.16)$$

where $W_{\parallel}(1)$ is the wake for the monopole mode.

To usefully compare the monopole and dipole modes a ratio is in fact needed. The magnitude of the combination of the two dipole modes can be found using Pythagoras's rule. The difference between the operating mode and the dipole mode in dBc are:

$$dBc = \log \left(\frac{\sqrt{X^2 + Y^2}}{M} \right) \cdot 10 \quad (5.17)$$

5.2.2 Results

As there is a random element to the dipole mode level the simulations must be repeated to find an average, as well as a worst and best case value. The properties of the modes are taken from the cavity study (see chap. 2) and so are applicable to the klystron design proposed (see table. 5.1).

Table 5.1: Properties and settings used for stability analysis

(a) Basic parameters

Property		Value
Number of beams	N_b	20
Beam current spread	Δ_b	5%
Number of repetitions		5000

(b) Mode properties

Operating Mode		Original Dipole Mode		New Dipole Mode	
Frequency	1.30 GHz	Frequency	1.35 GHz	Frequency	1.425 GHz
R/Q	170 Ω	R/Q	270 Ω	R/Q	270 Ω
Q	8000	Q	6000	Q	6000

5.2.2.1 Frequency of Dipole Mode

Although it is clear that by preference the dipole mode must be separated as far as possible from the operating mode and its harmonic frequencies, a measure of how far is not immediately available. If the bunch repetition frequency is held constant and the *mode* frequency altered (without any change in R/Q or Q_0) the effect on eq. 5.17 of dipole mode frequency can be investigated. It can be seen that by moving the cavity dipole mode there is an approximately 8dB extra drop in dipole excitation (see fig. 5.2).

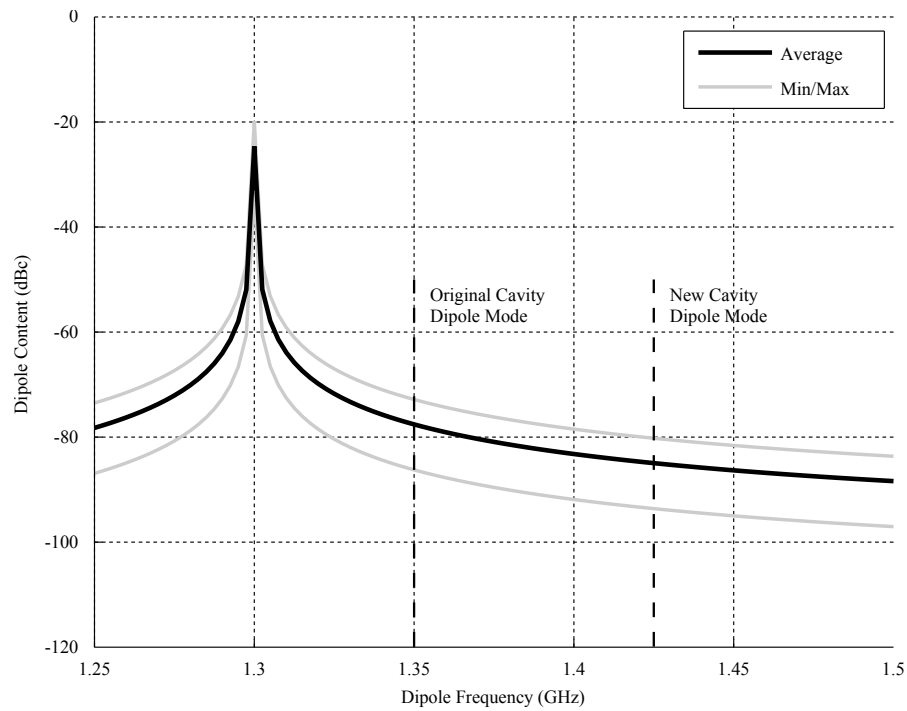


Figure 5.2: Position of appropriate dipole modes (denoted by dot-dashed lines) and their relative strength of interaction with the beam

5.2.2.2 Dipole Q factor

In order to lower the strength of interaction from the original dipole mode, intuitively an attempt would be made to damp the Q factor. However in addition to increasing the energy lost from the mode (and so the strength of interaction), this also increases the bandwidth of the mode and so increases the frequency range over which the mode will couple with the beam. Shown below is the effect of changing Q_0 on the dipole content of the beam for the original and new dipole mode of the coaxial cavity (see fig. 5.3). It can clearly be seen that the effect of the increase in the bandwidth of the mode outweighs the advantage of lowering the interaction strength.

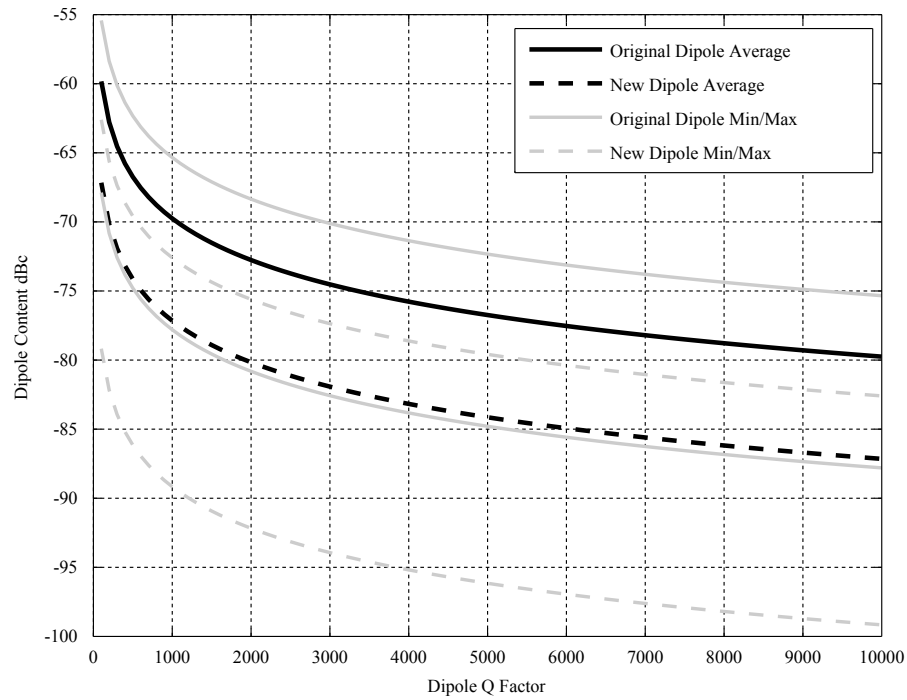


Figure 5.3: The effect of the dipole Q_0 on the dipole mode content for the original and new dipole modes. Showing the increasing dipole content with *decreasing* dipole Q_0 .

However if the Q_0 is damped to very low levels the damping effect dominates and the interaction decreases (see fig. 5.4). This effect is observed because the dipole mode is far enough away from the operating mode that it does not strongly interact when it has a high Q_0 , but close enough that the increase in bandwidth allows interaction.

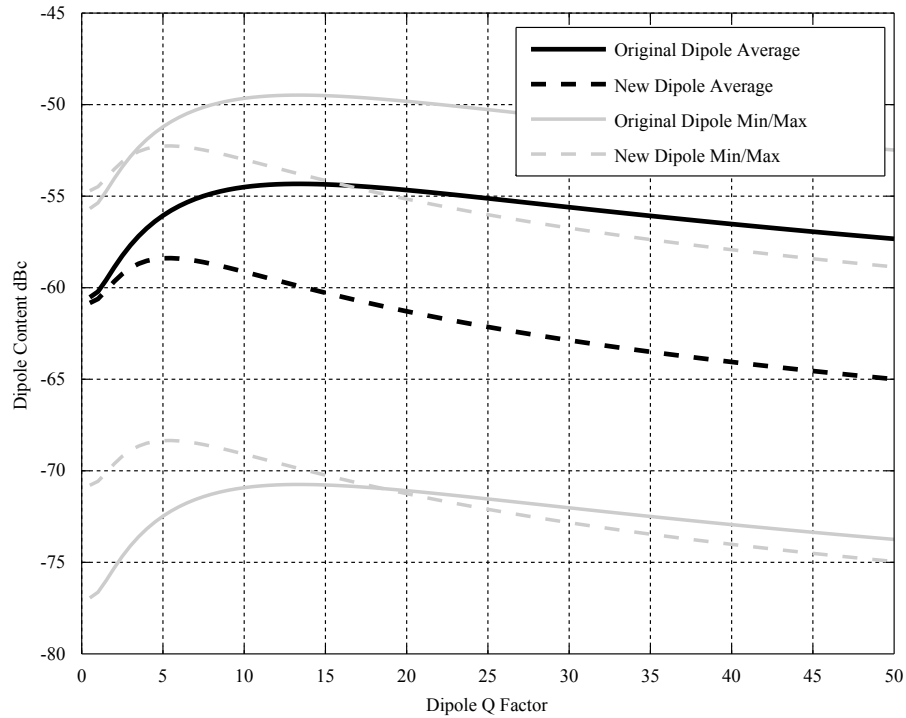


Figure 5.4: The effect of the dipole Q_0 on the dipole mode content for the original and new dipole modes. Showing the decreasing dipole content for very small Q_0 s.

5.2.2.3 Effect of Beam Current (I_b) Variation

As noted, perfectly identical beam currents are impractical leading to dipole mode excitation. With current technology a spread of 5% is expected, although the advantages of being able to reduce this are currently not clear.

It can be seen in fig. 5.5 that even a very small current variation excites an average dipole content of around -100 dBc (for the new dipole mode). Although the increase between 5% and 50% is a relatively modest 10 dBc. The dipole content of -240 dBc at 0% I_b variation can be attributed to numerical error in the calculation, this in reality should be $-\infty$ dBc.

If the I_b variation could be reduced to, for instance 1%, then a further ≈ 20 dBc drop in dipole content could be achieved. While this is perhaps not required for standard operation, once beams have been lost this could prove to be critical (see sec. 5.2.2.4).

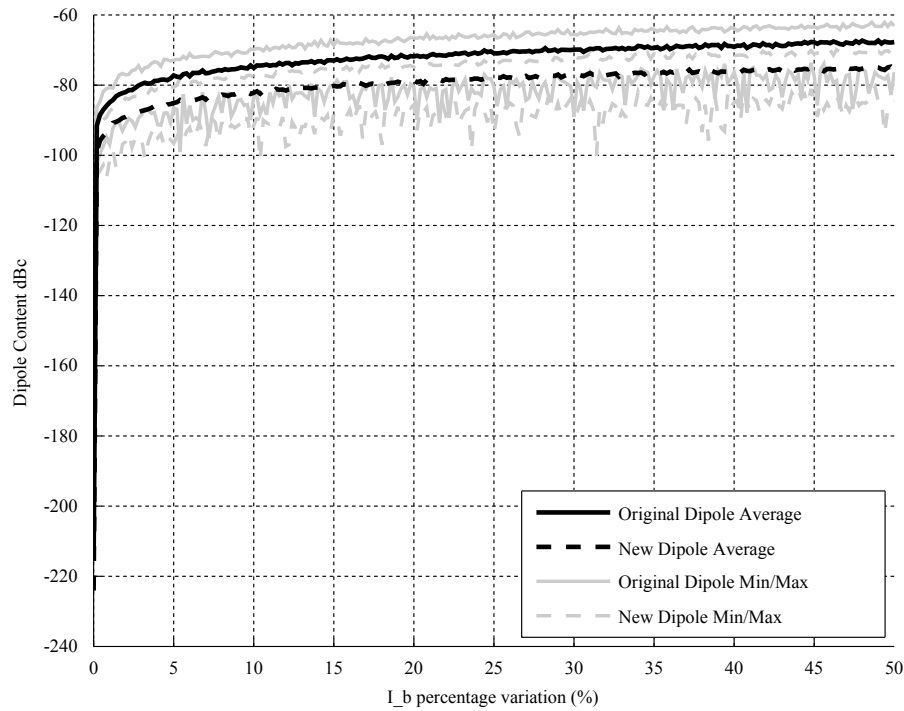


Figure 5.5: The effect of the variation in I_b between the beams. Showing that even a small variation creates an appreciable dipole component.

5.2.2.4 Loss of Beams

As the life of a klystron gun is finite even without the development of a fault, beams can be expected to be lost. The impact of the loss of a beam on a MBK is of interest if it goes beyond the proportional loss of power. The worst case, with respect to instability from beam loss, is if a group of beams were to be lost rather than the loss being distributed around the tube. It is this case which is investigated here.

It can be seen that the dipole content increases by ~ 15 dBc with the loss of just one beam. The impact beyond this is less severe but an increase of ~ 22 dBc is predicted for the loss of two beams and ~ 31 dBc for the loss of three (see fig. 5.6).

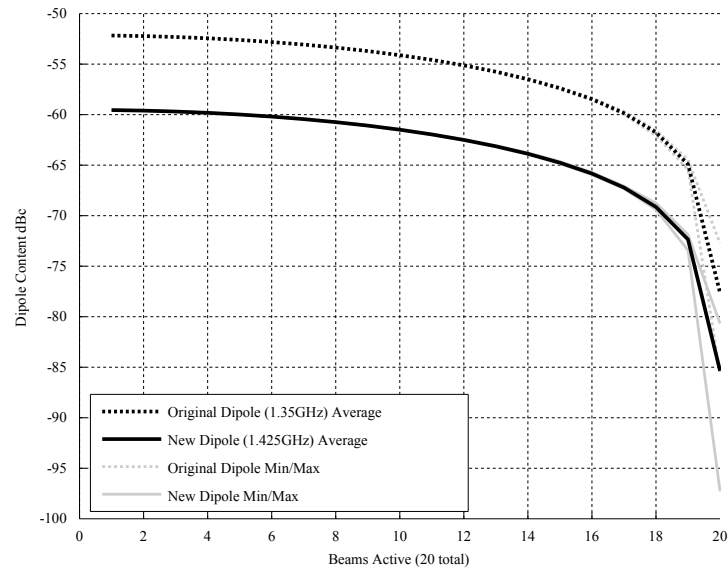


Figure 5.6: Effect of the loss of beams on the dipole content. The loss of just a single beam has a significant effect on dipole content. This effect decreases as number of beams lost increases.

It is intuitive that one beam would give the highest dipole content but in this case it is still low at 52.5 dBc. If the dipole mode were to have a frequency of 1.3 GHz (the operating frequency) its dipole content increases to a positive value, i.e. a strong dipole mode than monopole mode (see fig. 5.7).

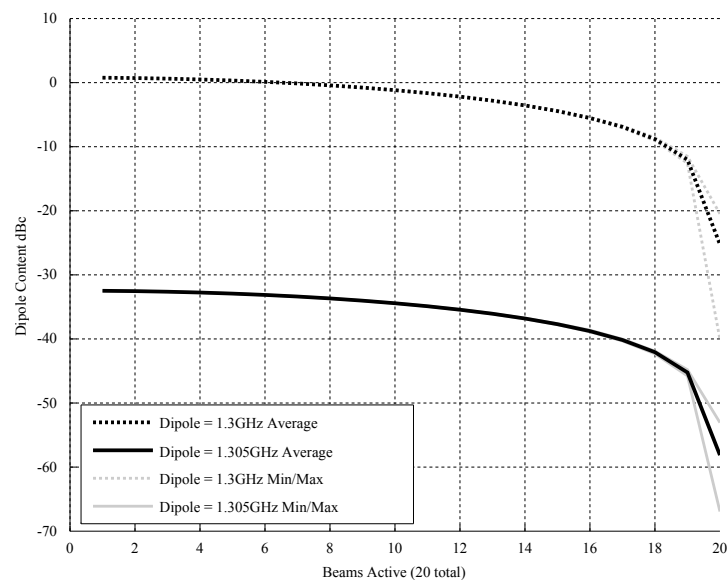


Figure 5.7: Effect of the loss of beams on the dipole content. For a dipole frequency of 1.3 GHz the high R/Q leads to a positive dipole content where only one beam is operational.

5.3 Conclusions

In this chapter a model of an output cavity of a multiple beam klystron represented as a series of independent gaps has been constructed using standard wake-field theory. Many complex effects seen in practical MBK are neglected for simplicity. Gain through the structure, transient effects and space-charge based cross talk between the beams have not been included in this model.

Intuitively it would be expected that lowering the Q_0 of a mode would reduce the interaction with the beam, this has been shown not to be the case. The broadening of the bandwidth outweighs the advantage of lowering the interaction strength and the dipole component actually increases. A much better solution is to change the mode's frequency.

Shifting the dipole mode by 75 MHz yields approximately 8 dBc extra drop in dipole excitation.

Thales reported 5% current matching capabilities introduce a natural -80 dBc and -85 dBc dipole component for the original and new dipole frequencies. The worst current variation investigated (50%) incurs a dipole component of -75 dBc and -70 dBc. The model fails to correctly predict that for 0% I_b variation the dipole content should be $-\infty$ dBc, but -240 dBc is very low.

If the I_b variation could be reduced to 1% then a further 20 dBc drop in dipole content could be achieved. While this is perhaps not required for standard operation, once beams have been lost this could prove to be critical.

The loss of a beam has the largest effect on dipole content as might be expected. Loss of a single beam can increase the dipole mode content to \sim -75 dBc (an increase of \sim 15 dBc) which when combined with the gain of a klystron could lead to instability. The impact of additional beams beyond this is less severe with a further increase of \sim 5 dBc and then \sim 4 dBc for the loss of two then three beams (see sec. 5.2.2.4).

The significance of the dipole mode content at these low levels can only be fully understood when combined with a model to take into account the gain of

the klystron. A dipole mode content of -50 dBc at the second cavity, excitation in the input cavity is unlikely, could lead to a much larger value at the output cavity. Further work could be undertaken to quantify this effect, and this model is only applicable to bunched beams which is not the case for the intermediate cavities. At this stage it appears that the dipole mode is in fact not a problem.

Chapter 6

Interaction Structure

6.1 Introduction

The properties of the interaction structure define the performance of the tube. It can be loosely described as the region the beam passes through after the electron gun and before the collector, excluding any coupling or HOM damping structures.

In modern klystrons it comprises 5-7 cavities with strictly defined properties although as few as two cavities is possible.

Many properties of the klystron can be calculated from simple analytical or empirical formulae. The properties of the cavities (as precise as they must be for the structure to work) have no clear relationship to the klystron specification.

In early work on klystrons it was suggested that a cavity spacing of $\frac{1}{4}\lambda_q$ would yield the greatest bunching (see fig. 6.1). As the beam is naturally bunched at $\frac{1}{4}\lambda_q$ it would appear that a cavity at this position would continue the bunching process more effectively than one before or after.[2] However with the aid of klystron interaction simulation tools and other refinements in klystron design it has been found that greater bunching still can be produced by placing the cavities slightly closer together. This also has the advantage of increasing the compactness of the tubes.

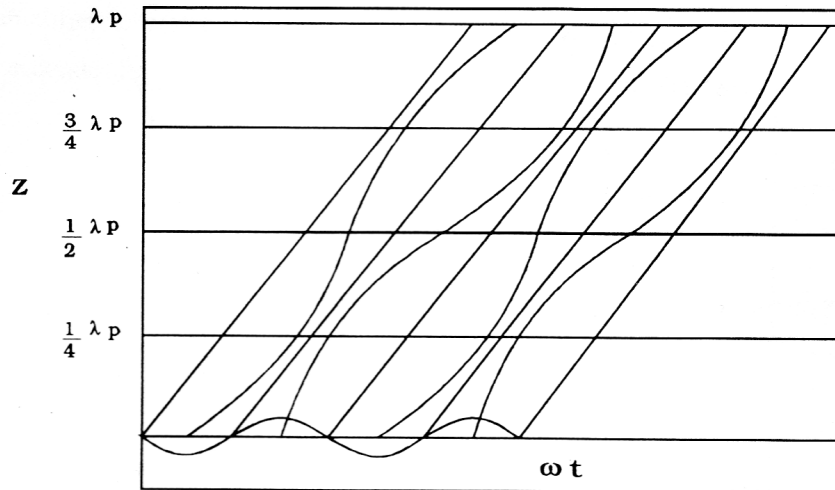


Figure 6.1: Applegate diagram showing the perceived advantage of $\frac{1}{4} \lambda_q$ cavity spacing for at least the first cavity[2]

Today high power klystrons are used almost exclusively by the scientific community, although lower powered klystrons still have a role in military radar and medical applications. Their use in broadcast systems have been almost completely replaced by solid state devices or IOTs (Inductive Output Tubes).

This change in role has altered the optimisation requirements. Broadcasters required a wide bandwidth to allow their system to adequately transmit data. Scientific applications often do not have this requirement. A small bandwidth of the order of 4 MHz @ 1dB is generally sufficient. This allows the klystron designer to concentrate their efforts on optimising efficiency, a much more important parameter for the scientific customer. The klystron is a common choice for the RF power supply of large devices. The power drawn by the klystron can be a large percentage of the power used by the experiment (for instance in CLIC) and the cost of the electricity a large percentage of the cost of running the machine. This has led to pressure for high efficiency klystrons with a narrow bandwidth requirement.

6.2 Manual Optimisation

As the possible cavity properties are now known for chapter (see chap. 2), an interaction structure can be calculated. The purpose of this chapter is not to create a fully optimised structure, but show that it is indeed possible to achieve an efficiency of 80% with the cavities and klystron parameters chosen, and go some way to justifying the use of the empirical relationship between perveance and efficiency. In order to simulate the interaction structure of a klystron under full load, a large signal code is required. This is significantly more complex and time consuming than the small signal code used for investigating tube bandwidths whose assumptions only hold when the input powers are very low.

Klys2D is a 2.5D PIC¹ klystron interaction structure simulation code². This software is used at Thales to simulate all their klystron designs and so has been extensively validated against experiment. It has the capability to simulate klystrons with and without radial particle movement due to focusing field and space charge effects. The initial simulations are often performed without radial movement in order to reduce the computing time needed, as by experience, these effects alter the klystron performance by a relatively small (although still important) amount.

Klys2D represents a klystron by its main parameters; gun voltage; beam current; beam and tunnel size; input power and focusing field (if applicable). In addition to this the cavities are characterised by their resonant frequency; gap length; drift length (from the previous cavity); R/Q and Q. This simple representation and the fact that the cavities have the same gap length throughout the different types means that they can be spit into two distinct groups.

- High R/Q ($\sim 200\Omega$)
 - Recessed re-entrant cavity
 - Fundamental mode coaxial cavity

¹particle in cell

²Kindly provided by Thales Electron Devices

- Low R/Q ($\sim 80\Omega$)
 - HOM Coaxial cavity
 - Whispering Gallery mode cavity

This allows the four plausible cavity types to be represented using only two interaction structures. The structures to be investigated will be of a conventional form comprising of 6 cavities:

- Input Cavity
- Gain Cavity
- 2nd Harmonic Cavity
- 2 Buncher Cavities
- Output Cavity (single gap)

Klys2D however does not have MBK support, the simulations are performed on a single beam and the input/output power is multiplied by the number of beams. This limits the usefulness of Klys2D for investigating the stability of any given interaction structure. The R/Q quoted in the cavity optimisation chapter (see chap. 2) is suitable for these simulations as it has already been adjusted to take into account the distribution of stored energy between the beams.

The target of the interaction structure simulations is to find a structure which displays the target gain; efficiency; output power and bandwidth. The structures were optimised by starting with a guessed structure and then, for each cavity in turn, performing a sweep over the cavity frequency and the drift length *after* that cavity. This allows the difference in bunching to be seen at the output of the klystron as an increase or decrease in output power. After the cavity property optimisation the input and output couplings are chosen using a parameter sweep.

After this has been performed twice the klystron is optimised for output power/efficiency only and not bandwidth. This is far from an exhaustive optimisation, relying purely on brute force but the structures should give an idea of what is possible if more time and experience was applied to the problem.

6.2.1 High R/Q Interaction Structures

6.2.1.1 Initial Parameters

Using the cavity and klystron properties used for the initial investigation a suitable interaction structure was found (see fig. 6.2). The first four harmonics of the modulation current, normalised with respect to the DC beam current I_{b0} , are plotted against distance along the tube. The effect of the 2nd harmonic cavity (no. 3) on the 2nd harmonic of the modulation current is clearly visible in the trace. The sharp drop of all harmonics after the output cavity is indicative of efficient power take off.

This design almost achieves the design target of 80% efficiency, but it did clearly illustrate some flaws in the klystron specification. The gap field in the output cavity for instance is higher than the 10 kV/mm figure used by Thales as a safe value (see table. 6.1)[78]. The tightness of the bunch achievable with the high R/Q structure is clearly displayed using an Applegate diagram (see fig. 6.3). The interaction structure is 1.8 m long which will call for a long solenoid, which also due to the r_{pcr} , will have a large diameter which could cause problems with the focusing system design. The 14 W input power equates to a 300 W input signal power to drive a full 20 beam MBK based on this structure. The input Q_e is 200 and output external Q_e is 55, these are easily achievable couplings and should not cause problems with their design.

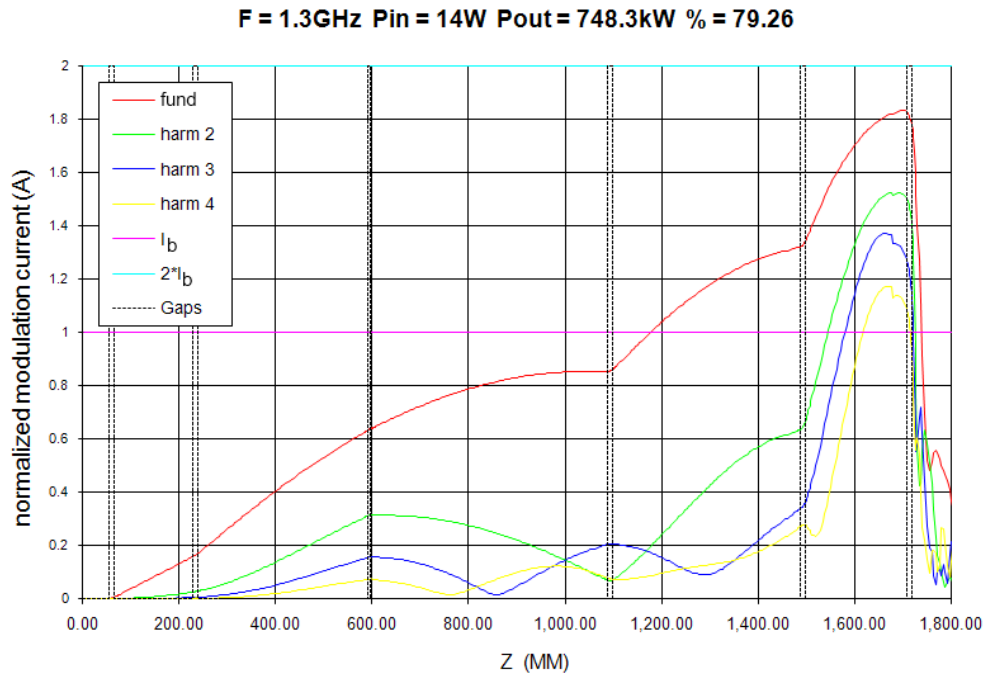


Figure 6.2: ~80% efficient high R/Q (Recessed re-entrant and fundamental mode coaxial cavity) interaction structure at a power of 748 kW per beam (14.9 MW in total) with 14 W input power

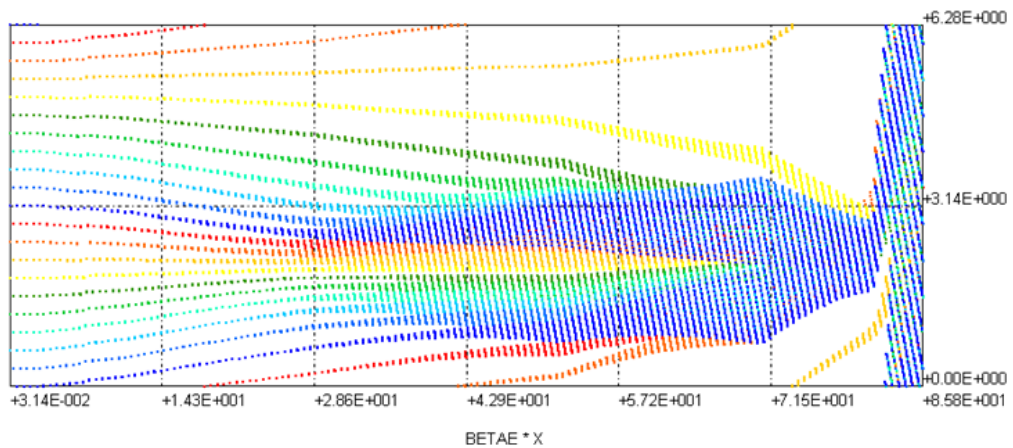


Figure 6.3: High R/Q Applegate diagram relating to the above interaction structure (see fig. 6.2)

The limit of electric field in a klystron gap used by Thales is defined as 10 kV/mm. This interaction structure has a output gap voltage of 113.3 kV which is a field of 10.69 kV/mm, above the limit (see table. 6.1).

Table 6.1: High R/Q cavity gap voltages

Gap	Voltage (kV)	kV/mm
1	6.647	0.63
2	4.532	0.43
3	0.3788	0.06
4	14.68	1.38
5	44.62	4.21
6	113.3	10.69 (>10 kV/mm)

6.2.1.2 Revised Parameters

The revision of the parameters used some lessons learnt from Thales' experience and the initial interaction simulation³. The new interaction structure has much longer gaps which should aid beam-cavity interaction and lower the output gap voltage, although it will lower the frequencies which have a danger of monotron oscillation.

An error in data input was made at this stage and the beam current was entered as 8.21 A, compared to the correct value of 8.15A. This leads to a slightly higher perveance and beam power. An 80% efficiency now gives an output power of 755 kW, compared to the correct power of 750 kW. The total output power of the klystron is now 15100 kW and the microperveance 2.11, from 2.09. As the error is small and the problems repeating this work significant this incorrect current is used in this section and the evolutionary algorithm chapter to allow comparison.

With these changes and using the initial structure as a starting point for the optimisation, the new interaction structure again reaches an efficiency of 80%. The input power for this interaction structure per beam is now lower at 10 W or 200 W for a full MBK. This structure however is much longer at around 2.2m long (from 1.8m), exacerbating the focusing problems that there may be. The slight kinks on the harmonic plots are indicative of a klystron on the cusp of reflecting electrons, a situation that must be avoided in simulation to ensure

³Private discussion with Thales Electron Devices Technical Staff

that reflection is at a minimum in a practical tube to ensure stability.[78]

The first drift length is much longer and the gradient of the first harmonic much lower than for the initial parameters. This may be an artifact of the optimisation and unnecessary or indicative of the lower R/Q used. The input coupling has also increased by 50% to $Q_e = 300$ with the output coupling remaining at $Q_e = 60$ (see fig. 6.4). Although the high efficiency suggests good bunching, the Applegate diagram indicates that it is not as tightly bunched, although this could be an artifact of the different number of "particles" in each of the diagrams (see fig. 6.5).

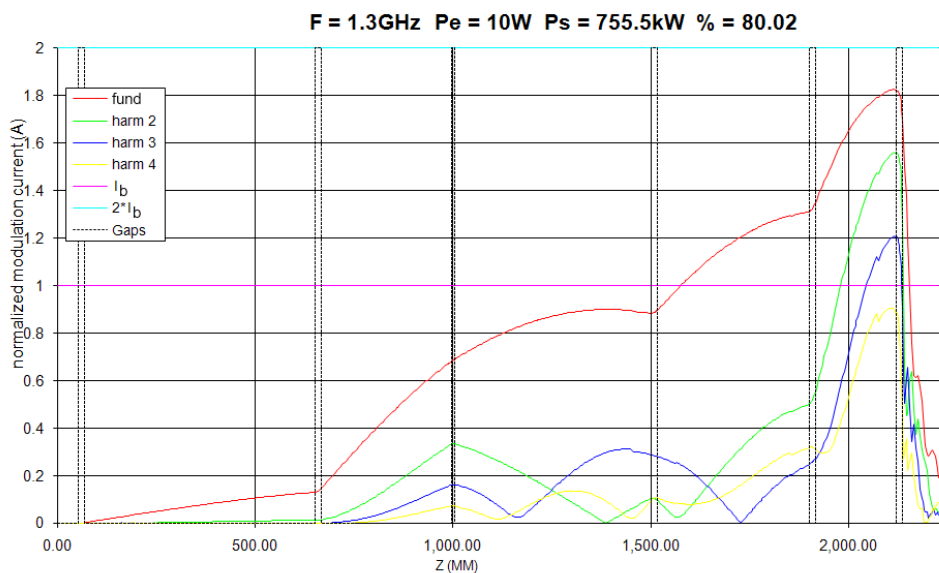


Figure 6.4: ~80% efficient high R/Q (Recessed re-entrant and fundamental mode coaxial cavity) interaction structure using the revised parameters at a power of 755 kW per beam (15.1 MW in total) with 10 W input power

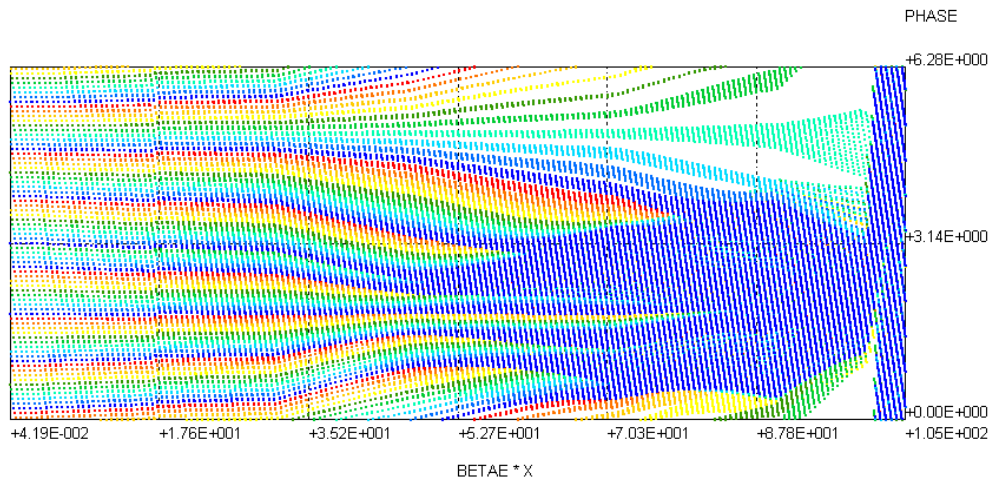


Figure 6.5: High R/Q Applegate diagram relating to the above interaction structure using revised parameters (see fig. 6.4)

Table 6.2: High R/Q revised klystron parameters cavity gap voltages

	Klys2D (kV)	kV/mm
1	1.8	0.11
2	11.8	0.74
3	0.386	0.04
4	13.6	0.85
5	36.7	2.29
6	116.6	7.29

The table of cavity voltages shows that the increase in gap length has served to lower the gap voltage per mm to an acceptable level (7.29 kV/mm from 10.69 kV/mm) even though the gap voltage is in fact slightly higher (116.6 kV from 113.3) (see table. 6.2). This parity between output gap voltages and output power neatly illustrates the relationship between them. The input voltage for both initial and revised parameters is 115 kV, similar to the output gap voltage as would be expected. The gap field must of course slow down and almost stop the electrons at the gap without reflecting them. With a voltage too low the electrons could retain too much of their energy; with a voltage too high electrons would be reflected back up the tube.

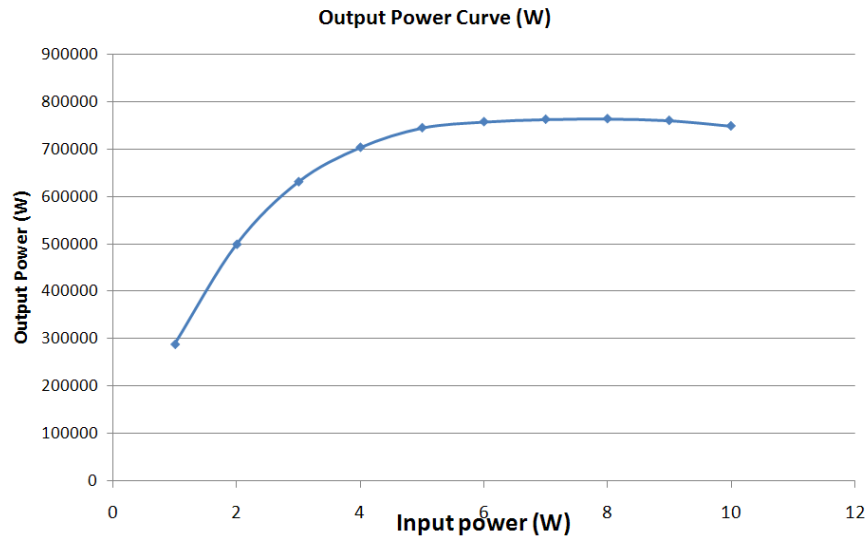


Figure 6.6: Interaction structure transfer curve for a klystron using the high R/Q cavities with the revised parameters

The transfer curve of this structure shows that this klystron would be running very much in the saturated region of the graph (see fig. 6.6). Ideally a klystron of this type would run with as low an input power as possible, perhaps 6 W in this case. This means that this structure is in all probability not the optimal design.

6.2.1.3 Magnetic Field

The simulation was repeated but this time radial movement and a magnetic field were introduced. The structure was not optimised using these simulations, as they are included as a validation for the omission of radial movement, indicated by the experience of Thales in both use of their code and klystron design. Although this inclusion affected the simulation complexity, the output results did not greatly differ from the simulations where it was omitted. The output power is reduced to 748 kW and the efficiency to 79.23%.

The only major change required is an increased beam pipe around the output cavity for reasons clear from the drift paths of the bunches (see fig. 6.8). Conceptually this can be easily explained by the increased longitudinal bunching encouraged around the output cavity increasing the space charge forces

and so expanding the beam radially, which from simulation, requires an increase in drift tube radius. Scalloping can clearly be seen as the beam moves along the tube, even before the bunches have formed.

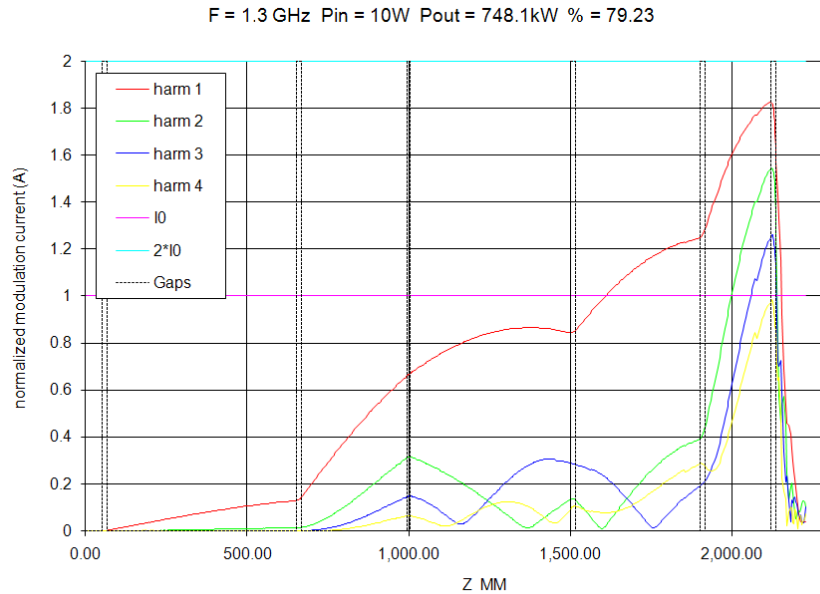


Figure 6.7: ~80% efficient high R/Q (Recessed re-entrant and fundamental mode coaxial cavity) interaction structure using the revised parameters at a power of 755 kW per beam (14.9 MW in total) with 10 W input power with efficiency and radial movement and magnetic field simulated

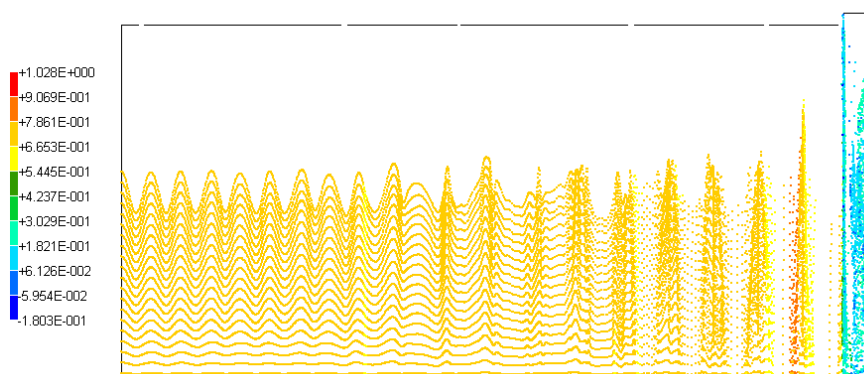


Figure 6.8: Particle tracks down the above klystron interaction structure with radial movement and magnetic field simulated showing the need for increased drift tube radius

6.2.2 Low R/Q Interaction Structures

6.2.2.1 Initial Parameters

The low R/Q structure requires high input and output Q_e s in order to encourage modulation and to couple power out of the tube due to their low R/Q. The external Q factors of the input and output cavities are 715 and 200 respectively compared to 200 and 55 for the high R/Q structure (see sec. 6.2.1). Although the input coupling is particularly high for a klystron and could cause problems in design, Thales' experience and evidence from other cavity applications suggests that such a coupler could be made⁴.

The 80% efficiency expected is not found in this optimisation, again ignoring radial movement and without a focusing field. An efficiency of just 72% and power of 677 kW is achieved (13.5 MW in total). This poor performance can only be due to the lower R/Q of the cavities, as this is the only parameter to have changed.

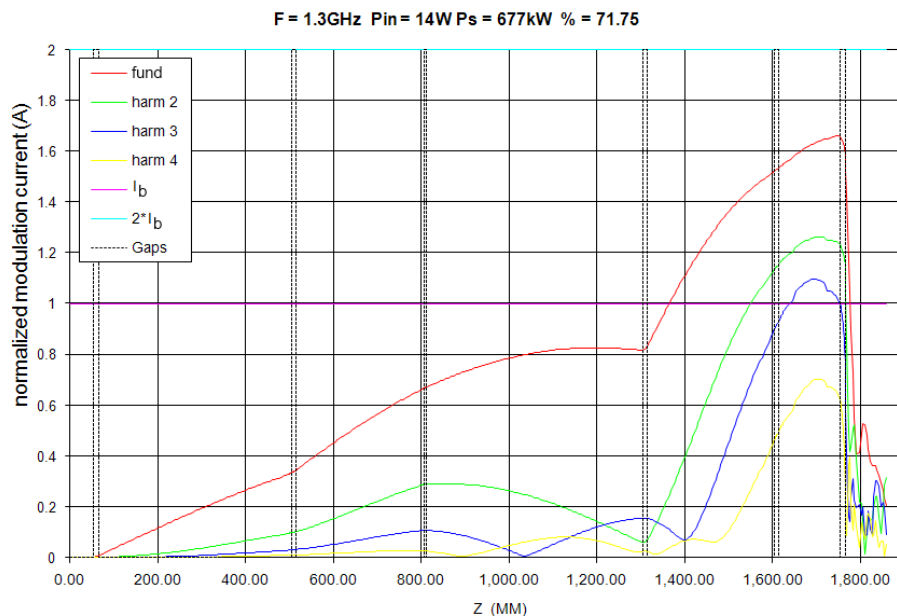


Figure 6.9: ~70% efficient low R/Q (HOM coaxial and whispering gallery cavity) interaction structure at a power of 677 kW per beam (13.5 MW in total) with 14 W input power

⁴Private discussion with Thales Electron Devices Technical Staff

A clear difference in the interaction structure (between low and high R/Q) is the distance between the first and second cavities (see fig. 6.2). Even with the high input Q the modulation is small, as evidenced by the lower gradient of the first harmonic of the modulation current. The bunches are given more time to form before being weakly bunched by the second cavity. It is the lack of interaction with the second cavity which leads to the poor performance of the tube.

Strangely the 4th cavity is best placed after the modulation maxima where the space charge has begun to spread the bunches as evidenced by a reduction of 1st harmonic current (see fig. 6.9).

The optimisation of this structure is much more complex leading to an inefficient optimisation process when compared to the high R/Q structure. The efficiency found should not be taken as a maximum, but it can be seen as indicative of the difficulty in optimising this structure when compared to the high R/Q structure, as a similar length of time was spent working on both. The 5th cavity in this structure for instance does not appear to affect the modulation current of the beam in any of the harmonics. Effectively this is a 5 cavity klystron but altering the properties of this 5th cavity only serves to lower the efficiency. The 6th and final cavity also suffers from weak interaction with the beams, removing less of the modulation from the beam than the high R/Q structure (see fig. 6.8).

The length of this interaction structure is comparable to the high R/Q structure at 1.8m so although the beams interact weakly with the cavities this does not have a large effect on klystron size as might be expected.

6.3 Conclusion

This chapter contains three interaction structures optimised by hand, one for the low R/Q cavities (HOM coaxial and whispering gallery mode) and two for the high R/Q cavities; using initial and revised parameters (recessed re-entrant

and fundamental mode coaxial).

From the initial parameters it is clear to see that the high R/Q structure is preferable to the low R/Q structure. Critically the target 80% theoretical efficiency was achieved with relative ease with a high R/Q. Although there is no guarantee these structures are a true optimum, a similar level of effort has provided markedly different results. Additionally the input and output Q_e are both conservatively valued, simplifying the design of their associated couplers.

The extreme electric field at the output gap illustrates the problem of insufficient gap length. This led to a revision of the initial parameters to alleviate this and other issues (see sec. 2.9).

The revised parameters led to a maintained efficiency and output power with a reduced modulation power of 10 W compared to 14 W. The input coupling has increased by 50% but still remains in a practical region. The voltage at the output gap has increased from the initial parameters but the extended gap lowers the field to an acceptable value. The lower interaction with the cavities, a direct result of the lower R/Q, leads to a slightly longer interaction structure.

The structure found is operating in the saturated region of its transfer curve, demonstrating that this would not be used as a final design, however it does suggest that such a design could be possible. If the input power is reduced below saturation then the efficiency of the tube falls to 5-10% below the target efficiency.

A further simulation where the particles are allowed to move radially, focused by the focusing field, shows only a small difference from the simulation without focusing.

The main conclusion, the success of the high R/Q cavity apart, is that manual optimisation of klystron interaction structures is unnecessarily time consuming. This task could perhaps be performed more efficiently with the use of an appropriate algorithm.

Chapter 7

Literature Review Evolutionary

Algorithm

7.1 Introduction

As noted in section 6.3 manual optimisation is a very slow process with no guarantee of a true optimal result. A human operator can only reasonably deal with 2 or 3 input variables at once and more than one objective is difficult without making use of a technique such as gradient descent which is susceptible to becoming trapping in local minima. When confronted with a 20 dimensional problem one approach is to break it down into a series of 2 dimensional problems and perform an iterative optimisation. As it is hard to deal with multiple objectives simultaneously a non optimum solution will result. Use of a cost function¹ can help define a preference between values but even if a trade off is successfully made, a single solution will be found, as discussed in this chapter a series of solutions may have advantages.

Another major drawback of manual optimisation, apart from the quality of the optimum, can be defined in terms of duty cycle. As manual optimisation requires a human operator to analyse the data; identify a new set of simulations and initialise it, there is a period of “downtime” where the computational

¹a mathematical relationship used to define the relative quality of a result

resources are idle.

There are analytical and empirical relationships between many of the properties of a klystron as discussed in section 1.5. However there are some ill defined decisions that must be made before a klystron is constructed:

- Cavity frequencies
- Drift lengths between cavities
- External coupling of the input and output cavities

A manual optimisation of these parameters is possible but consideration of much more than output power is difficult. Unfortunately klystron specifications and operation mandate the use of several objectives:

- High Efficiency, output power, gain (equivalent for a constant input power)
- Short length of interaction
- Specification bandwidth
- No reflected electrons

This justifies the approach taken by tube manufacturers, where they adapt and update older tube designs to meet new specifications. Replacing the human factor in the decision making process leads to an increase in the duty cycle and decrease in time taken for optimisation.

The largest flaw in the automated approach is the loss of the intelligent decision making process. The evolutionary algorithm approach attempts to sidestep this problem, not only through an element of brute force, but by analysing the solution space. A purpose-made algorithm is always faster and more efficient, but a genetic/evolutionary algorithm has the advantages of wide flexibility and well-known behaviour.

7.2 What is an EA (Evolutionary Algorithm)?

Evolutionary algorithms are an example of a probabilistic, parallel, meta-heuristic search technique inspired by natural evolution. Unlike many optimisation methods, EAs work from a population of solutions sampling the search space to characterise the solution space. Recombination operators are applied to share information, and mutation operators to explore new regions of the search space. This allows for complex, epistatic (non-linear) and multi-modal (multiple optima and/or sub-optima) optimisation problems to be efficiently explored.[79, 14]

Real-coded EAs (RCEA) are used in a wide range of scientific applications and their characteristics are well understood. They are well suited to performing optimisations in problems with a high dimensionality.

Basic genetic algorithms encode the decision space in binary, using the analogy of gene manipulation, and then cut and splice these, often using the concept of mutation to add an extra random element. Their operation is analogous to evolution; successful given time but unfocused.[79]

RCEAs are distinct from genetic algorithms in that they manipulate the parameters numerically rather than encoding them as binary information. Using the available (often incomplete) information, various strategies are used to analyse the solution space to identify good areas worthy of further investigation.[79, 18]

The basis of the algorithm is two multi-dimensional spaces; the decision space comprises vectors containing the input variables; the solution space comprises vectors (or scalars) of the output variables or objectives. Collectively they are known as the search space.[79]

The fitness of any individual is a scalar measure of how optimal it is. It should allow individuals to be compared correctly. For a single solution maximisation problem this can simply be the solution, bigger is better. For a multi-objective (multiple solution) problem this is more complicated and will be explained in greater depth. The basic theorem of fitness is that as fitness increases

the individuals become closer and closer to the optimum.

The representation of the search space as known by the algorithm is described using points known as individuals consisting of: [14, 79]

Individual $\{x, z, f, \gamma\} \in \mathbf{X}$ - the decision x and solution z vectors (with the solution often a scalar), the fitness f and the ranking score γ

Population $\{\mathbf{X}_{(1)}, \mathbf{X}_{(2)}, \mathbf{X}_{(3)}, \dots, \mathbf{X}_{(|P|)}\} \in \mathbf{P}$ - contains the individuals in a population

The individuals provide the EAs only view of the problem in their mapping of the decision x to solution z .

A generic EA manipulates the individuals using a series of general operator types:

Evaluation $\mathbf{F}(x) = z$ - evaluate the individuals, converting the decision vector x to a solution vector z

Fitness $[f](z) = f$ - convert the solution vector z to a scalar fitness f

Selection $\mathbf{P}_p^{(g)} = [\omega](\mathbf{P}^{(g)})$ - using some method to analyse the population and select the most appropriate for the next operator.

Recombination $\mathbf{P}_\phi^{(g+1)} = [R](\mathbf{P}_p^{(g)})$ - combine the parents in some way to produce offspring

Mutation $x' = [\mu](x)$ - apply some kind of (often random) transformation to the decision vectors

The most basic algorithm structure is illustrated in fig. 7.1. The initial population is generated, with the decision space populated at random within certain problem-specific constraints. This decision space is mapped to the solution space using the simulation or problem that is to be optimised. This solution space is analysed to select the best individuals which are then used in reproduction (or recombination). These are then calculated ready to undergo selection once more. This loop continues until some kind of stopping condition is met.

Each of these steps can be undertaken using many different approaches which will be discussed in this chapter.

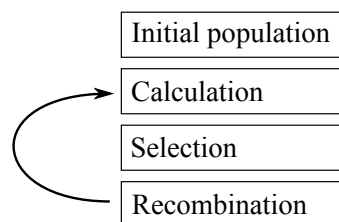


Figure 7.1: Basic Evolutionary Algorithm structure

7.3 Selection

Selection is based on the fundamental assumption that good individuals when combined will produce good offspring and that there is a correlation between parent and offspring fitness.[14] This comes from a concept basic to biological evolution, that of heritability without which selection is futile at best.

With this in mind selections aim to improve the average quality of the population by focusing on promising areas of the search space.[14] Most selection schemes consciously only deal with the exclusively scalar fitness of individuals, having delegated generation of this fitness to the fitness operator $[f]$, allowing the use of very general selection methods. However it does show that in fact the fitness operator $[f]$ must ensure that an individual's quality is measured using fitness f . [14, 80]

Selection is partly based on mathematical analysis, partly inspired by biological systems. The choice of selection parameters (and selection method) defines the relationship between exploitation and exploration. If the search pressure is too high the algorithm will converge prematurely. If it is too low, although it will explore the search space effectively, it will never converge on a solution.[14]

There are many ways to benchmark a selection method. For instance takeover time is the number of generations needed, in the absence of mutation

and reproduction, for the population to become completely filled by a single individual.[14] Selection intensity, suggested by Mühlenbein and Schlierkamp-Voosen [81], describes the change in the average fitness of the population due to selection:[14]

$$I = \frac{\bar{M}^* - \bar{M}}{\bar{\sigma}} \quad (7.1)$$

where $\bar{\sigma}$ is the fitness variance; \bar{M} is the average fitness and \bar{M}^* is the average fitness after selection.

Blickle and Thiele [14] attempt to compare a range of selection operators using the fitness distribution before and after selection has taken place. The full analysis is beyond the scope of this review but the findings are of interest.

The conversion from a score to a selected individual is performed by a notional roulette wheel. Each individual in the pool is represented on the wheel by a region proportional in size to its score.[79] The drawback of this approach is that the wheel must be spun once per selection. Although tournament selection does not make use of the roulette wheel, it also requires N tournaments.[18]

In “stochastic universal sampling” (SUS) where there are N individuals to be selected, N evenly spaced selection marks are used. As there are N marks, all N individuals are chosen at once.[79]

Much of this section is concerned in fact with ranking operators which convert fitness into a score. There are three commonly used mechanisms for performing the actual selection from this score after ranking has taken place [79]:

$[\omega]_R$ Uniform Random - each individual has an equal chance of being selected

$[\omega]_{RW}$ Roulette Wheel - each individual has a chance of being selected directly proportional to their score. This is analogous to regions on a roulette wheel proportional to the individual’s score. In the programming domain this can be implemented using a uniform random number generator and predefined ranges.[18]

$[\omega]_{USS}$ Universal Stochastic Sampling - the same roulette wheel as above now has N evenly spaced selection pointers (where N is the number to be selected). So one turn of the wheel selects N individuals. Blickle and Thiele [14] consider this to be an optimal sampling algorithm for use with ranking. This however does change the selection properties of the ranking operators as the selection of each individual is related due to the even spacing of the marks. This helps to maintain diversity but other techniques can be used to achieve a similar result. SUS is not applicable to tournament selection.[14]

7.3.1 Tournament Selection

Algorithm 7.1 Tournament selection algorithm [14]

input population \mathbf{P} with fitness distribution f and T tournament size

$[\omega]_R(\mathbf{P}) = \mathbf{P}_t$ where $\mathbf{P}_t \subset \mathbf{P}$, $|\mathbf{P}_t| = T$ selected randomly with fitness distribution f_t
 index $j = \max_i(f_t)$ (index of individual with max f)

output $\mathbf{X}_{t(j)}$

Tournament selection is a true selection operator with no reliance on the application $[\omega]_{RW}$ or $[\omega]_{SUS}$ to the output. A predetermined number of individuals are selected at random from the population using $[\omega]_R(\mathbf{P})$ and placed into the tournament pool \mathbf{P}_t , which takes part in a tournament; with the individual with the highest fitness in \mathbf{P}_t selected and returned. Often a binary tournament ($|\mathbf{P}_t|$ of 2) is used, although any size is possible. The size of the tournament T strongly affects the selection intensity, with a larger tournament increasing the chances of less fit individuals being beaten, and so increasing the selection intensity (see fig. 7.2a) and altering the probability distribution (see fig. 7.2b). [14]

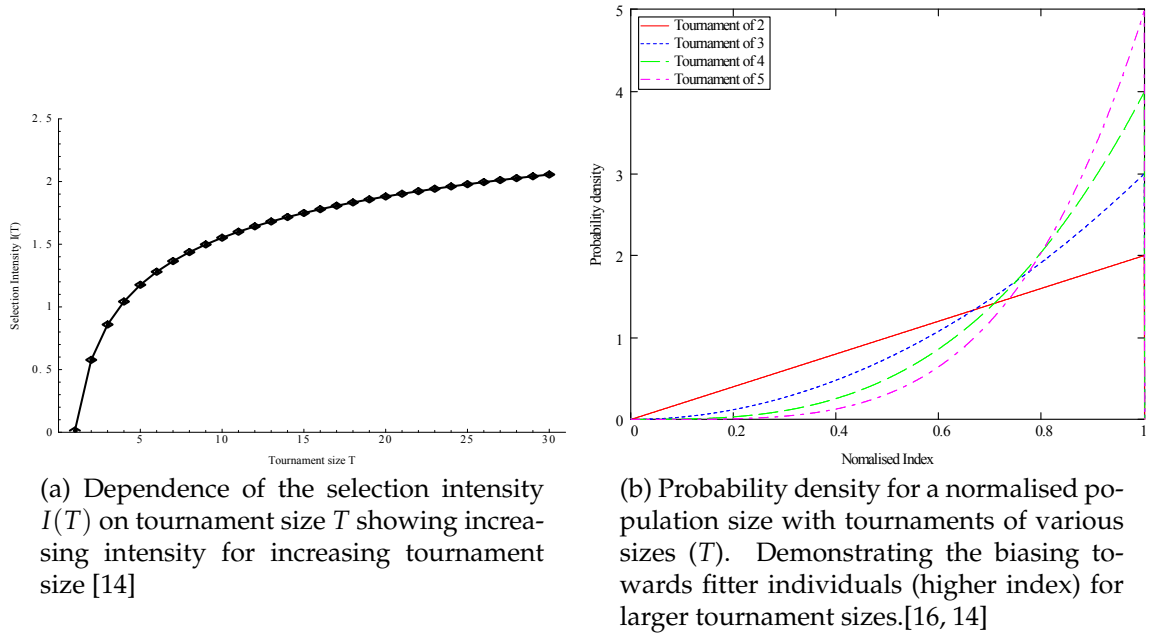


Figure 7.2: Tournament selection

7.3.2 Truncation Selection

Algorithm 7.2 Truncation selection algorithm [14]

input population \mathbf{P} with fitness distribution f and t truncation factor

$f_{t(i)} = f_{(i)}$ where $i \in \{(1-t)|\mathbf{P}|, \dots, |\mathbf{P}|\}$ where \mathbf{P} (and so f) is sorted with ascending $f_{(i)}$

$[\omega]_R \mathbf{P}_t \mapsto \mathbf{X}_{t(n)}$ randomly select an individual from the population

output $\mathbf{X}_{t(n)}$

Truncation selection also does not assign a score. It first generates a pool of the $(1 - T)|\mathbf{P}|$ (where T is the truncation factor and $|\mathbf{P}|$ the size of the population) best individuals (by fitness). It then selects from this pool with equal probability. Using this technique ensures that the worse fraction of the population can never take part in reproduction, but inevitably genetic material is discarded.[14]

As the truncation factor decreases, more of the population is discarded, the selection intensity increases (see fig. 7.3a) and the probability distribution is ske-

wed towards fitter individuals (see fig. 7.3b).

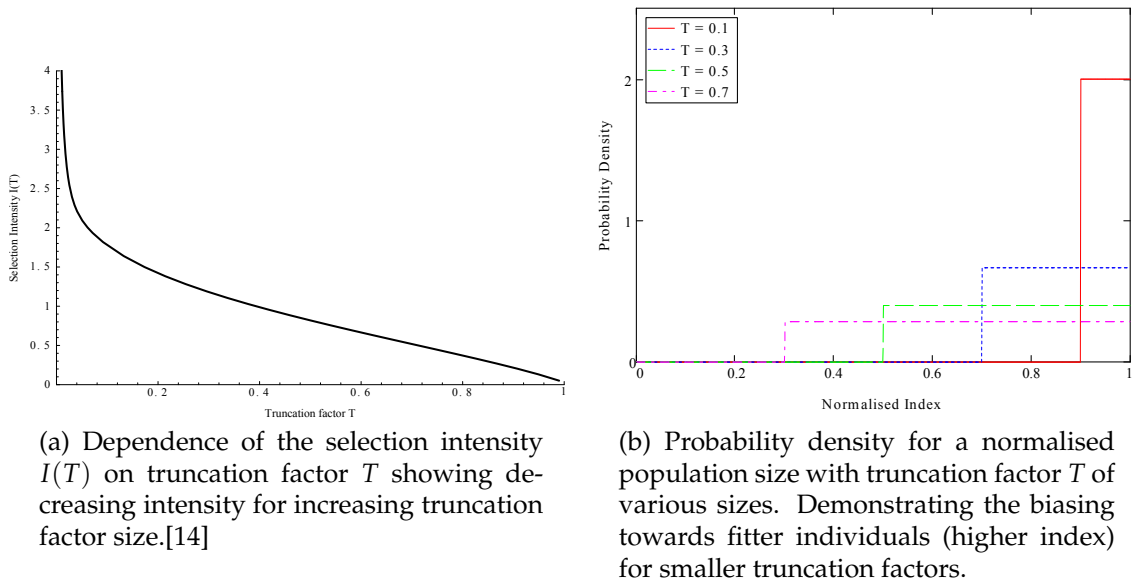


Figure 7.3: Truncation selection

7.3.3 Linear Ranking

Algorithm 7.3 Linear ranking selection algorithm [14]

input population \mathbf{P} with fitness distribution f and parameters η^+ and η^-

$$\gamma = \{\text{eq. 7.2}\}$$

$[\omega]_{RW}(\mathbf{P}) = \mathbf{X}_{(i)}$ select $\mathbf{X}_{(i)}$ using roulette wheel with a weighting of γ

output $\mathbf{X}_{(i)}$

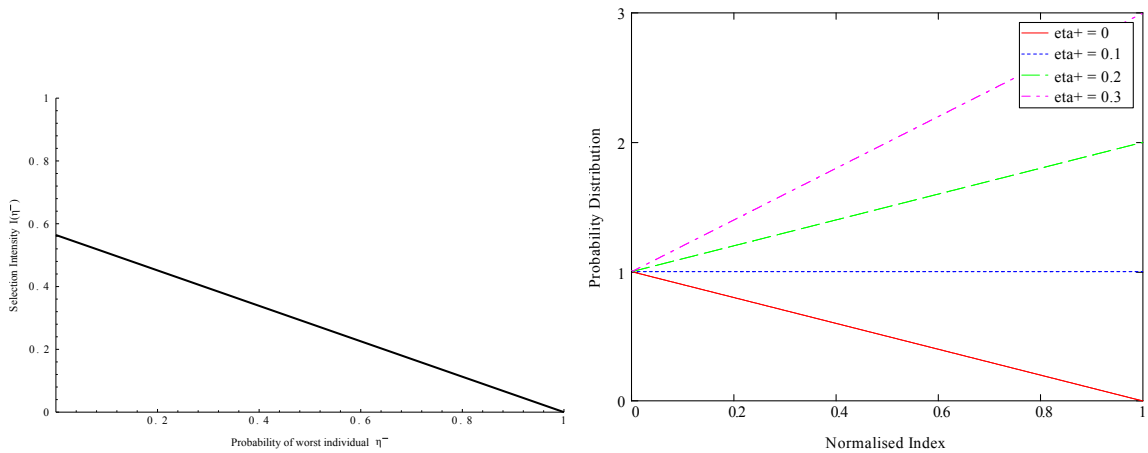
Linear ranking masks the underlying fitness data, eliminating potential serious disadvantages of fitness proportional selection. Where an individual with a very large fitness may in fact dominate selection to such a degree that no other individuals are chosen. Although the fitness is used during an initial sort the probability of selection is purely related to the position in the population or index. Each position i in the population has an associated probability or “score”. To generate this score a simple linear function is used to translate the position i in the sorted population of f into a score $\gamma_{(i)}$. [14]

$$\gamma_{(i)} = \frac{1}{|\mathbf{P}|} \left(\eta^- + (\eta^+ - \eta^-) \frac{i-1}{|\mathbf{P}|-1} \right) \tag{7.2}$$

$$i \in \{1, \dots, |\mathbf{P}|\} \tag{7.3}$$

where $\eta^- / |\mathbf{P}|$ is the score of the worst individual and $\eta^+ / |\mathbf{P}|$ the score of the best.[14]

It can be shown by experiment [80] and mathematically [14] that a binary tournament performs identically to linear ranking, but is advantageous because it does not require a sort of the population. Increasing the probability of selecting the worst individual decreases the search intensity (see fig. 7.4a) as expected, and increasing the probability of selecting the best individual biases the search towards them.[14]



(a) Dependence of the selection intensity $I(\eta^-)$ on probability of worst individual η^- . Showing decreasing intensity for increasing probability.[14]

(b) Probability density for a normalised population size with variable probability of best individual selection η^+ . Demonstrating the biasing towards fitter individuals (higher index) for higher probabilities (as expected).

Figure 7.4: Linear ranking

7.3.4 Exponential Ranking

Algorithm 7.4 Exponential ranking selection algorithm [14]

input population \mathbf{P} with fitness distribution f and parameter c

$\gamma = \text{eq. 7.4}$

$[\omega]_{RW}(\mathbf{P}) = \mathbf{X}_{(i)}$ select $\mathbf{X}_{(i)}$ using roulette wheel with a weighting of γ

output $\mathbf{X}_{(i)}$

Exponential ranking is similar in spirit to linear ranking. Again the population is sorted and the position i in the population is used to score the individuals. This time, instead of a linear function constrained by two parameters, an exponential function defined by a single parameter is used (see eq. 7.4). The parameter $0 < c < 1$, the closer c is to 1 the lower the “exponentiality” of the selection:[14]

$$\gamma(i) = \frac{c^{|\mathbf{P}|-i}}{\sum_{j=1}^{|\mathbf{P}|} c^{|\mathbf{P}|-j}} \quad (7.4)$$

$$i \in \{1, \dots, |\mathbf{P}|\} \quad (7.5)$$

The lower the value of c the higher the selection intensity (see fig. 7.5a) and the more biased towards fitter individuals the selection becomes (see fig. 7.5b).[14]

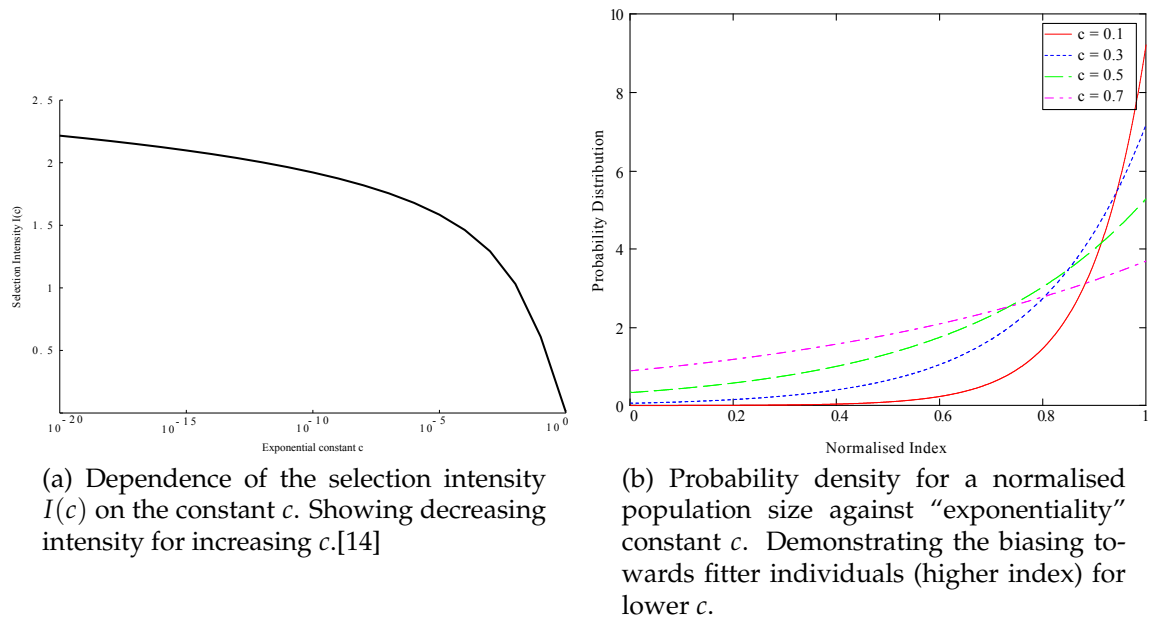


Figure 7.5: Exponential ranking

7.3.5 Loss of Diversity

The loss of diversity is the proportion of individuals which are not selected and their "genetic information lost". This can effect the algorithm's ability to search the space effectively, for instance in the case of multiple optima. [14] Bickel and Thiele [14] found truncation has a higher loss of diversity for an arbitrary fitness distribution than any other selection method tested when compared with a fixed selection intensity (see fig. 7.6). The remaining operators perform similarly in this regard.

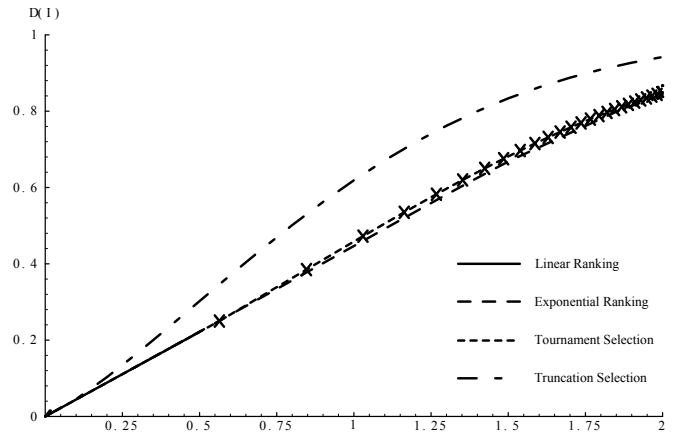


Figure 7.6: Dependence on the loss of diversity on selection intensity for Linear Ranking, Exponential Ranking and Tournament and Truncation selection. Blickle and Thiele [14]

7.4 Recombination

The recombination operator is where a GA and RCEA differ most. Often a recombination operator is known as a crossover operator in reference to those used in GAs where the binary coded data are spliced together to spread the genetic material. In a RCEA it is instead mathematical analysis which creates the new decision vectors. This allows much more complex analysis of the space.[15]

For these purposes only parent based operators will be discussed. That is those recombination operators which act over a population of parents \mathbf{P}_p selected using a method described above (see sec. 7.3). Sometimes a particular parent (\mathbf{x}_p) is nominated to be the mother ($\mathbf{x}_{p(m)}$), this takes special meaning in some reproduction methods.[15]

The UNDX (Unimodal Normal Distribution Crossover) operator uses multiple parents and generates offspring in a distribution around the centre of mass of the parent population in an approach known as mean-centric recombination (see fig. 7.7a). [15]

The SPX (Simplex Crossover) operator, on the other hand, generates offspring within a region bounded by the parents with equal probability (see fig. 7.7b).[15] UNDX creates children tightly close to the centre of the parents, all of which

may have similar behaviour good or bad.[15] Higuchi et al. [82] found that SPX performs better than UNDX. This said, the test relied on the MGG model, (see alg. 7.7) where the recombination operator creates many offspring (all of which must be, potentially expensively, evaluated) and discards the worst, and SPX required a much larger offspring pool than UNDX.

Selection is based on the fundamental assumption that good individuals when combined will produce good offspring [14]. There is however there is no reason to suspect that the centroid of the parents would have a high fitness, especially if they are sparse in the decision space. [15]

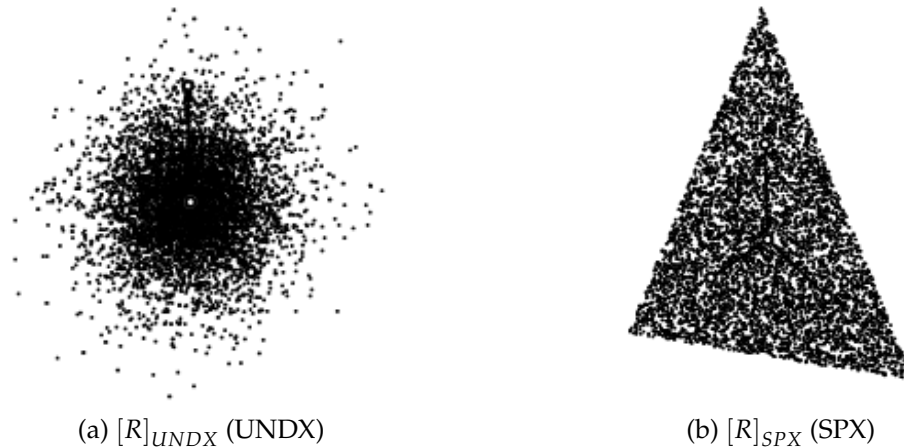
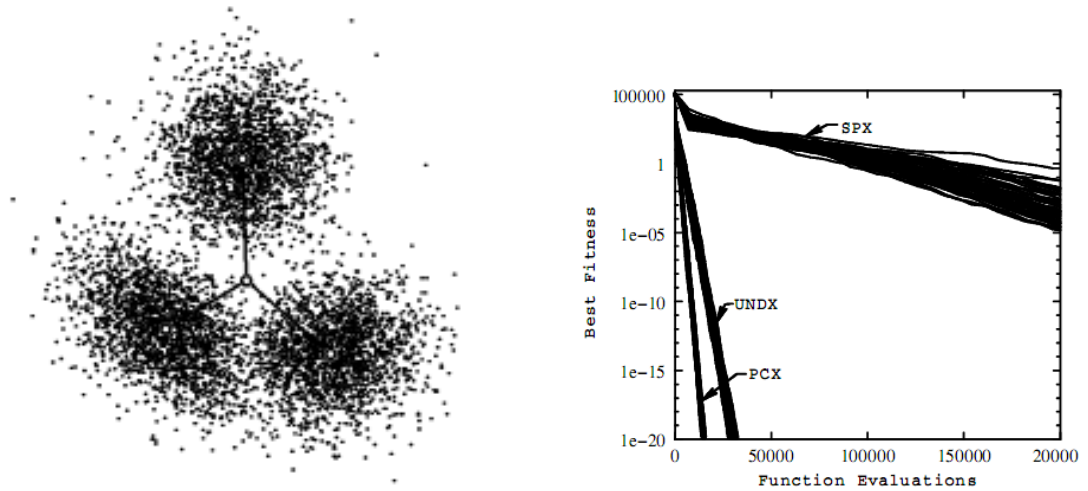


Figure 7.7: Probability distributions of UNDX and SPX recombination operators

7.4.1 PCX (Parent Centric Recombination)

Another approach would be to assign to each parent an equal probability of creating an offspring nearby (see fig. 7.8a). This approach is called parent-centric recombination. The use of a mean-centric probability distribution (such as that used in UNDX) have been found not to be as efficient as a parent-centric approach (see fig. 7.8b). PCX also uses the mother paradigm as a basis for calculation.[15]



(a) Probability distributions of $[R]_{PCX}$ (PCX) recombination operators[15]

(b) Best fitness against function evaluations. Showing that PCX's better prediction leads to better convergence. [15]

Figure 7.8: Properties of PCX[15]

Before Deb [15] it was unclear that a mean-centric or parent centric approach was best. Deb defines the parent-centric crossover² (PCX) operator in [15] as follows (with updated nomenclature):

“The SBX operator assigns a higher probability for an offspring to remain closer to the parents than away from parents. We use this parent-centric concept and modify the UNDX operator as follows. The mean vector $\bar{\mathbf{x}}_p$ of the chosen p (p is taken as $N + 1$ in this document) parents is computed. For each offspring, one parent $\mathbf{x}_{p(m)}$ is chosen with equal probability. The direction vector is calculated $\mathbf{d}_m = \mathbf{x}_{p(m)} - \bar{\mathbf{x}}_p$. Thereafter, from each of the other $(p - 1)$ parents, perpendicular distances d_p to the line \mathbf{d}_m are computed and their average \bar{d}_p is found. The offspring is created as follows:

$$\mathbf{x}_\phi = \mathbf{x}_{p(m)} + \omega_\zeta \mathbf{d}_m + \sum_{i=1, i \neq m}^p \omega_\eta \bar{d}_p \hat{\mathbf{e}}_{(i)} \quad (7.6)$$

where $\hat{\mathbf{e}}$ are the $(p - 1)$ orthonormal bases that span the subspace

²the terms “crossover” and “recombination” are equivalent

perpendicular to \mathbf{d}_m .”[15]

In one dimension, this results in the probability distribution shown in fig. 7.9.

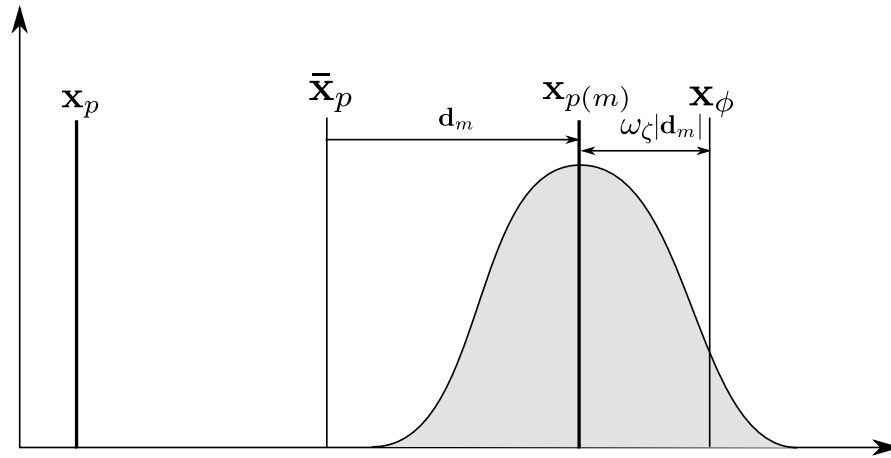


Figure 7.9: Probability distribution of offspring (\mathbf{x}_ϕ) generated by PCX in one dimension (where $\bar{\mathbf{x}}_p$ is the mean of the parents and $\mathbf{x}_{p(m)}$ is the mother)

To see the shortcomings of this method, this method must be considered in simplified terms:[15]

Algorithm 7.5 Basic PCX[15]

1. From the population \mathbf{P} , a sub population of p parents, \mathbf{P}_p are selected
 2. Of the parents (\mathbf{P}_p), one is selected at random to be the mother $\mathbf{x}_{p(m)}$
 3. The mean of the parents ($\bar{\mathbf{x}}_p$) is calculated
 4. The vector \mathbf{d}_m is defined as the vector from $\mathbf{x}_{p(m)}$ to $\bar{\mathbf{x}}_p$.
 5. The perpendicular distance from the line vector \mathbf{d}_m and each other parent is calculated, and the mean \bar{d}_p taken.
 6. A set of orthonormal bases (a set of perpendicular unit vectors) $\hat{\mathbf{e}}$, all perpendicular to \mathbf{d}_m are generated (these can not be defined specifically with the available information, but any set of orthonormal bases perpendicular to \mathbf{d}_m are acceptable).
 7. The offspring \mathbf{x}_ϕ is made equal to the mother $\mathbf{x}_{p(m)}$
 8. A normally distributed, 0 mean random magnitude ($\omega_\zeta \cdot |\mathbf{d}_m|$) vector is added in the direction \mathbf{d}_m standard deviation of $\sigma = \sigma_x |\mathbf{d}_m|$
 9. A normally distributed, 0 mean random magnitude ($\omega_\eta \cdot \bar{d}_p$) vector is added in each direction of $\hat{\mathbf{e}}$ with a standard deviation of $\sigma = \sigma_\eta \bar{d}_p$.
-

If a two dimensional search space is considered, it can be seen that PCX will search with a different power in two dimensions (σ_x in the direction \mathbf{d}_m ; and σ_η perpendicular) as shown in figure 7.10.

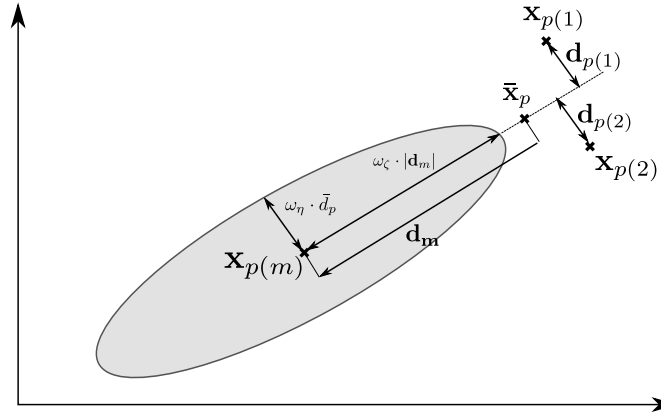


Figure 7.10: Probability contour at 1 standard deviation of offspring generated by PCX in two dimension with three parents ($\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$ and $\mathbf{x}^{(p)}$)
 $\bar{d}_p = \frac{d_{p(1)} + d_{p(2)}}{2}$ [15, 16]

The flaw in PCX lies in the fact that it bases its probability distribution on only two vectors. In a three or higher dimensional search space with dimensions of differing magnitudes this can cause PCX to search out of the allowable space.

The number of parents is strongly related to the dimensionality of the search space. It is important to represent fully each dimension when creating the probability distribution. A single parent for instance represents no dimensions as there is no distribution information. Two parents can represent one dimension as a line, three parents two dimensions and so on. Using this simple rule at least $N + 1$ (where N is the length of the decision vector) parents are needed to represent the space well.[15]

7.4.2 MrCX (Mother Centric Recombination)

As PCX defines its probability distribution from two vectors, Mother Centric Recombination (MrCX) bases it on N vectors (where N is the dimensionality of the decision space). This allows it to easily take into account differing magni-

tudes, trends and sensitivities in each dimension. In much the same way as PCX, MrCX defines orthonormal vectors of Gaussian distributed magnitude, oriented by the mother-to-mean-parent vector.[17]

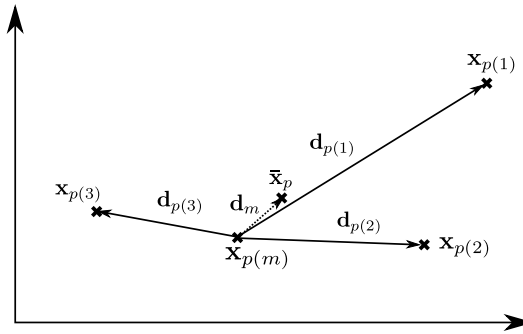
In this modified form, the offspring are generated as follows (see Fig. 7.11):

$$\mathbf{x}_\phi = \mathbf{x}_{p(m)} + \sigma_x \mathbf{v} \quad (7.7)$$

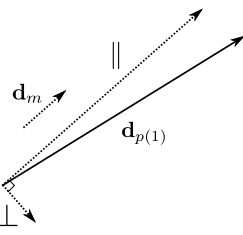
where \mathbf{v} is the MrCX vector defined below (see alg. 7.6). σ_x is a control parameter to be set by the user to set the search pressure. This results in the offspring being produced within a Gaussian distribution in N dimensions around the mother.

Algorithm 7.6 MrCX (bold text denotes alteration from PCX)[17]

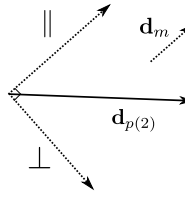
1. From the population \mathbf{P} , a sub population of p parents, \mathbf{P}_p are selected.
 2. Of the parents (\mathbf{P}_p), one is selected at random to be the mother $\mathbf{x}_{p(m)}$.
 3. The mean of the parents ($\bar{\mathbf{x}}_p$) is calculated.
 4. The vector \mathbf{d}_m is defined as the vector from $\mathbf{x}_{p(m)}$ to $\bar{\mathbf{x}}_p$.
 5. **The \perp and \parallel components w.r.t. \mathbf{d}_m of the distance ($\mathbf{d}_{p(i)}$) from $\mathbf{x}_{p(m)}$ to each parent (see figs. 7.11b, 7.11c and 7.11d).**
 6. **A set of orthonormal vectors \mathbf{e} , all \perp or \parallel to \mathbf{d}_m are generated by adding the \perp and \parallel components of $\mathbf{d}_{p(i)}$ in the same direction (see fig. 7.11e).**
 7. **The offspring \mathbf{x}_ϕ is made equal to the mother $\mathbf{x}_{p(m)}$.**
 8. **$\mathbf{v} = \frac{r\sigma_x}{p} \mathbf{e}$ where \mathbf{r} is a vector of Gaussian random numbers ($\bar{x} = 0, \sigma = 1$) (see fig. 7.11f)**
 9. **The offspring is then $\mathbf{x}_\phi = \mathbf{x}_\phi + \mathbf{v}$.**
-



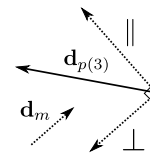
(a) Parent and mother distribution, showing offset vectors $\mathbf{d}_{p(i)}$ between parents and mother and \mathbf{d}_m between mother and the parent mean ($x_{p(i)}$ are parents, $x_{p(m)}$ is the mother, \bar{x}_p is the mean of the parents) (Steps 2-5 of Alg. 7.6)



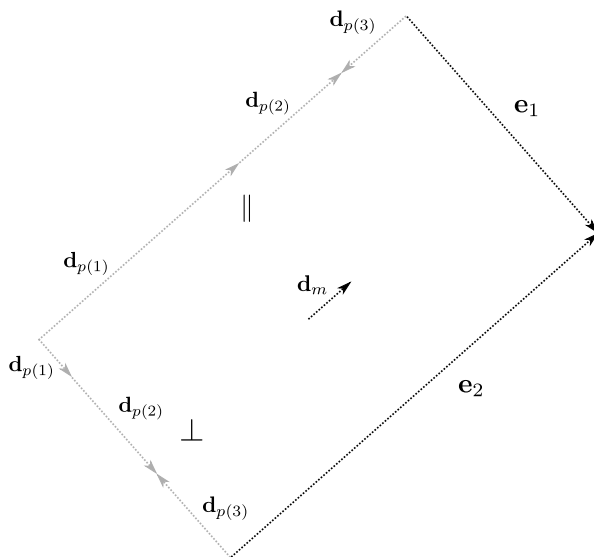
(b) $\mathbf{d}_{p(1)}$ split into components \parallel and \perp to \mathbf{d}_m (Step 5 of Alg. 7.6)



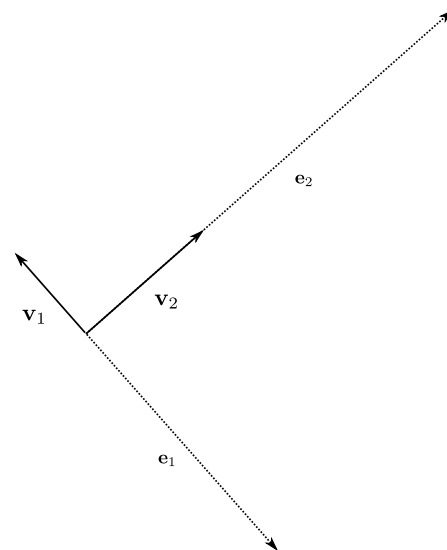
(c) $\mathbf{d}_{p(2)}$ split into components \parallel and \perp to \mathbf{d}_m (Step 5 of Alg. 7.6)



(d) $\mathbf{d}_{p(3)}$ split into components \parallel and \perp to \mathbf{d}_m (Step 5 of Alg. 7.6)



(e) Sum of \parallel and \perp components of $\mathbf{d}_{p(i)}$ (ignore -ve direction) to yield \mathbf{e} (Step 6 of Alg. 7.6)



(f) Multiply \mathbf{e} by a scaled Gaussian random number in each direction (Step 8 of Alg. 7.6)

Figure 7.11: Vector operations used in MrCX (to scale)[17](see alg. 7.6)

This approach solves the problem of biasing the reproduction probability distribution towards incorrect regions in high dimensional problems.

7.5 Population Management

Using the Minimal Generation Gap (MGG) model interweaves selection and reproduction. A pool of parents are selected and mated to provide offspring (which must be evaluated). Two randomly selected parents are replaced with the best and a randomly selected offspring (see alg. 7.7). The large offspring pool increases the chances of a good individual being found, retaining only the best compensates for a poor solution space prediction.[15]

Algorithm 7.7 Minimal Generation Gap (MGG) model[15]

input population $\mathbf{P}^{(g)}$ with fitness distribution f

1. $\mathbf{P}_p = [\omega]_R (\mathbf{P})$ select parents randomly from the population
2. $\mathbf{P}_\phi = [R] (\mathbf{P}_p)$ use recombination to generate offspring from parents
3. $\mathbf{P}_t = [\omega]_R (\mathbf{P})$ select two individuals at random from the population
4. Of \mathbf{P}_t , one is replaced in $\mathbf{P}^{(g)}$ with the best of offspring ($\mathbf{X}_{\phi(j)}$, $j = \max_i (f_\phi)$) and the other is replaced with a solution chosen by a roulette-wheel selection $[\omega]_{RW} \{ \mathbf{P}_\phi \cup \mathbf{P}_t \}$
5. $\mathbf{P}^{(g+1)} = \mathbf{P}^{(g)}$

output $\mathbf{P}^{(g+1)}$

Deb [15] found that replacing the roulette wheel with a block selection of the two best offspring modifies the MGG model to make it computationally faster. This method is called Generalised Generation Gap (G3) a steady state elite preserving algorithm for real parameter optimisation (see alg. 7.8). This method also acts to preserve elite solutions. [15]

Algorithm 7.8 Generalised Generation Gap (G3) model [15]

input population $\mathbf{P}^{(g)}$ with fitness distribution f

1. $\mathbf{P}_p = [\omega]_R \left\{ \mathbf{P} \cup \mathbf{X}_{(j)}, j = \max_i(f) \right\}$ where the tournament size is the size of the population resulting in the best being selected and the other parents randomly.
2. $\mathbf{P}_\phi = [R] (\mathbf{P}_p)$ use recombination to generate offspring from parents
3. $\mathbf{P}_t = [\omega]_R (\mathbf{P})$ select two parents at random from the population .
4. From $\mathbf{P}_\phi \cup (\mathbf{P}_t)$ choose the best two offspring and replace \mathbf{P}_t in $\mathbf{P}^{(g)}$ to make $\mathbf{P}^{(g+1)}$

output $\mathbf{P}^{(g+1)}$

If, for simplicity, a mutation operator is not used, a real-parameter recombination operator should satisfy the above two properties. UNDX and SPX notably do not fulfil both properties.[15] Beyer and Deb[15, 18] argued that a variation operator (a combination of the recombination and the mutation operator) should have the following two properties:

1. the population mean should remain unchanged
2. variance of the individual spacings must increase[15]

7.6 Multi-Objective EAs

In single objective optimisation the decision vector is optimised to give an optimal solution, which usually corresponds to a global maximum or minimum. However when there are multiple objectives to be optimised simultaneously such a simple approach is not possible.[19]

Often in a multi-objective optimisation it is not possible to simultaneously optimise all of the solutions. Such a situation is described as an objective conflict, how a multi-objective algorithm deals with such a situation is critical to its usefulness.[19]

There are many simple classical methods for optimisation which define a substitute problem, allowing the vector problem to be reduced to a scalar version which is optimisable using standard EA techniques.[19]

7.6.1 Classical Methods

A number of classical methods have been used to attempt to use optimisation algorithms for multi-objective problems. Their approaches range from converting the solution to a scalar, to dealing with the objectives separately.[19] Some of the approaches used will be outlined here, all of which are simpler than the specialised multi-objective approaches described shortly (see fig. 7.12).

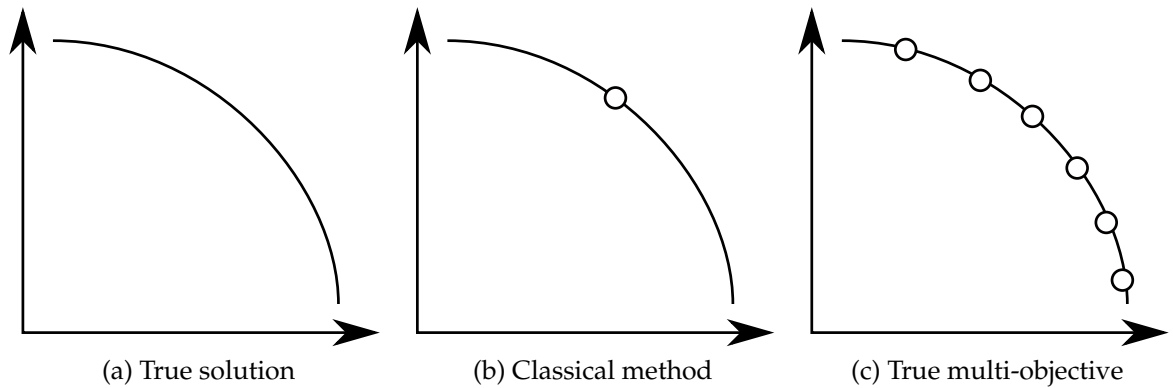


Figure 7.12: Representations of a multi-objective problem. Showing the advantage of a true multi-objective solution over a classical method.[18]

7.6.1.1 Method of Objective Weighting

This is the simplest possible method of converting the solution vector to a scalar. The elements are simply summed using a weighting vector \mathbf{w} to create a scalar fitness.

$$f_{(i)} = \sum_{j=0}^{|\mathbf{z}|} \mathbf{w}_{(i)} \cdot \mathbf{z}_{(i)}(j) \quad (7.8)$$

where $\mathbf{z}_{(j)}(n)$ is the j^{th} element of the i^{th} individuals solution.

A solution obtained with equal weights may mathematically offer the least objective conflict, it may not be a feasible (or indeed optimum) solution to a real world problem. In essence \mathbf{w} defines the point on the true solution which the algorithm will converge to (see fig. 7.12b). It should be noted that a poor selection of \mathbf{w} should not result in a non optimal point. As the curve is not produced by the algorithm, and will not be known, selection of \mathbf{w} is non-trivial in the general case.[19]

7.6.1.2 Method of Distance Functions

This method works by finding the distance (in solution space) from a predefined point known as the demand level vector \mathbf{y} . This point must be specified prior to running the algorithm from any information available. In this case if the demand level is set poorly a non optimal solution could be found with no indication given from the algorithm. [19]

$$f_{(i)} = \left[\sum_{j=1}^M |z_{(i)}(j) - y_j|^r \right]^{1/r} \quad (7.9)$$

often $r = 2$ is chosen with \mathbf{y} as individual optima of objectives [19, 83]

If the individual optima are not well known then a good value of \mathbf{y} is difficult to define. In a many-objective problem individual optima are non-trivial to find. Multi-modality will only further confuse the decision leading to a poor solution.[19]

7.6.1.3 Min-Max Formulation

Similarly to the method of distance functions, this method has a target fitness \mathbf{y} vector. In this method the normalised differences of each are calculated and the highest returned. This attempts to minimise the objective conflict, that is to say it minimises the occurrence of a individual which is very poor in any objectives. The advantage of this method is that \mathbf{y} is largely arbitrary as it is

only the difference which is taken into account. Poor selection can however cause problems.[19]

For a minimisation problem:

$$f_{(i)} = \max(Z_j(i)) \quad (7.10)$$

$$j \in \{1, 2, \dots, M\} \quad (7.11)$$

where $Z_j(i)$ is calculated for a target fitness vector $\mathbf{y}(j) > 0$ as follows:[19]

$$Z_j(i) = \frac{\mathbf{z}_{(i)}(j) - \mathbf{y}(j)}{\mathbf{y}(j)} \quad (7.12)$$

$$j \in \{1, 2, \dots, M\} \quad (7.13)$$

where $\mathbf{y}(j)$ is the j^{th} member of \mathbf{y} .

This method performs well with equally prioritised objectives. [19]

7.6.1.4 Drawbacks of Classical Methods

All of the methods described above require at least some knowledge of the problem to be solved, often in-depth knowledge of the solution space that may not be available or trivial to derive. Most importantly they all converge to a single point by design, which may not even be guaranteed to be an optimal solution. This point is often defined before the algorithm is started and strongly influences the solution found, moreover a non optimal solution is more likely when the problems solution space is badly defined causing the starting conditions to be badly chosen.[19]

None of the methods described are effective if some of the objectives are subject to noise or discontinuities in the solution space.[19] All of the classical methods are sensitive to large solution magnitude and range differences which, although they can be solved using weights and offsets, can cause some objec-

tives to be drowned out by others if care is not taken.[16]

7.6.2 Domination

In any multi-objective simulation there exists a set of solutions which are demonstrably superior to all other solutions. Such a set is known as a Pareto optimal or non-dominated set. The remaining solution space can be described as dominated. Critically no member of the Pareto optimal set can be said to be superior to any other. Also given the opportunity, a decision maker may be interested in knowing alternative solutions.[19]

In place of the classical methods non-dominated sorting is used (where \mathbf{a} and \mathbf{b} are solution vectors).

$$\mathbf{a} \preceq \mathbf{b} \quad (7.14)$$

where $\preceq =$ all components are greater than or equal to.

The effect of eq. (see eq. 7.14) on a solution space dominated by a series of points can be seen in fig. 7.13.

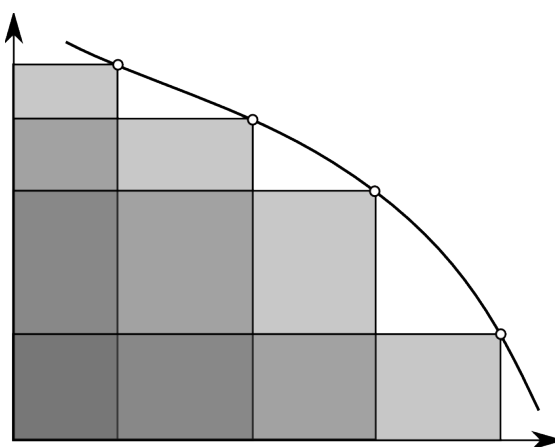


Figure 7.13: The effect of eq. (see eq. 7.14) on the space behind a series of points. The darker colours represent a more highly dominated area.

These vectors can be arranged into Pareto fronts (solution vectors where no one characteristic can be further optimised without detrimental effect on the

other characteristics).

These are named Pareto fronts because in practice there could be more than one Pareto optimal in the Pareto optimal set creating a front in the solution space.[19]

The concept of domination is common between the two current most successful multi-objective approaches NSGAII (see sec. 7.6.6) and SPEA2 (see sec. 7.6.8). These two approaches also attempt to preserve diversity along the Pareto fronts and reduce clumping.[20]

7.6.3 Clustering (k^{th} Nearest)

Density estimation can provide information on the formation of a Pareto front. Densely populated regions should be less fit than sparsely populated regions in order to investigate the full Pareto front, rather than just one section of it.[23] Zitzler et al. [23] use a technique called k^{th} nearest neighbour to perform this task.

The density at any point is defined as a function of the distance to the k^{th} nearest neighbouring point. To arrive at this value the distances between all of the individuals in solution space are calculated and sorted. The element in the k^{th} position is used as the distance to define the density $\mathbf{d}_{i,j}$. The function to define the density is:[23]

$$D(i) = \frac{1}{\mathbf{d}_{i,j} + 2} \quad (7.15)$$

where k is set to the square root of the sample size (in this case the size of the population(s)).

7.6.4 VEGA (Vector Evaluated Genetic Algorithm)

This multi objective approach notably does *not* rely mainly on the concept of domination. It performs selection individually on each of the objectives to fill

the mating pool. This neatly avoids the problem of differing objective magnitudes and tricky selection of weights although in some cases it is biased towards certain regions or individuals.[19]

Srinivas and Deb attempts to avoid biasing using two approaches, a wealth distribution scheme (which does use the concept of domination to penalise dominated individuals by limiting their numbers in the mating pool) and a cross breeding scheme (which mates an individual with another having the maximum euclidean distance in the solution space). The former fails when there are very few non-dominated individuals and Schaffer concluded the latter was not effective.[19, 84]

Algorithm 7.9 Basic VEGA algorithm neglecting wealth distribution and cross breeding [19]

1. Generate an initial population \mathbf{P}
 2. For each generation $g \in \{0, 1, \dots, G\}$
 - (a) Evaluate using $\mathbf{F}(\mathbf{x})$
 - (b) For each objective n
 - i. Select p/n individuals \mathbf{P}_p using objective n
 - (c) Reproduction
-

7.6.5 NSGA (Non-dominated Sorting Genetic Algorithm)

NSGA takes a different approach. Rather than altering the selection scheme then formulating additional techniques to address inadequacies, NSGA instead alters the allocation of fitness.[19]

NSGA can be seen as a fitness assignment technique. Each individual is ranked according to their level of non-domination. First all of the non-dominated solutions are found and given a high fitness equal for all the front. This fitness is divided by a number proportional to the number of individuals near by in

the solution space in a process known as sharing.

$$Share(d_{i,j}) = \begin{cases} 1 - \left(\frac{d_{i,j}}{\sigma_{share}}\right) & \text{if } d_{i,j} < \sigma_{share}; \\ 0 & \text{otherwise.} \end{cases} \quad (7.16)$$

where $d_{i,j}$ is the distance between two individuals \mathbf{X}_i and \mathbf{X}_j and σ_{share} is the maximum distance allowable between two individuals for this factor to apply. $Share(d_{i,j})$ is summed for \mathbf{X}_i for all \mathbf{X}_j where $j \in \{0, \dots, |\mathbf{P}|\}$ but $i \neq j$. [19]

This has the effect of penalising individuals who are tightly grouped in the solution space helping spread the solutions along the Pareto front. This first non-dominated front is then ignored and the next non-dominated front processed in a similar fashion. The high initial fitness is kept lower than the lowest fitness in the previous Pareto front ensuring each front is less desirable than the last. This method favours non-dominated solutions and tends toward Pareto-optimal fronts. [19]

By way of demonstration of the advantages of domination a test function used by Shaffer was simulated by Deb using VEGA and NSGA.

Shaffer's F1 test function is defined as follows:

$$f_{11} = x^2 \quad (7.17)$$

$$f_{12} = (x - 2)^2 \quad (7.18)$$

VEGA converges quickly to a fair solution (not shown) but then fixates on 3 points, loosing the utility of the solution. NSGA converges more slowly but ensures the Pareto front is well defined nicely demonstrating the advantages of the concept of domination (see fig. 7.14).

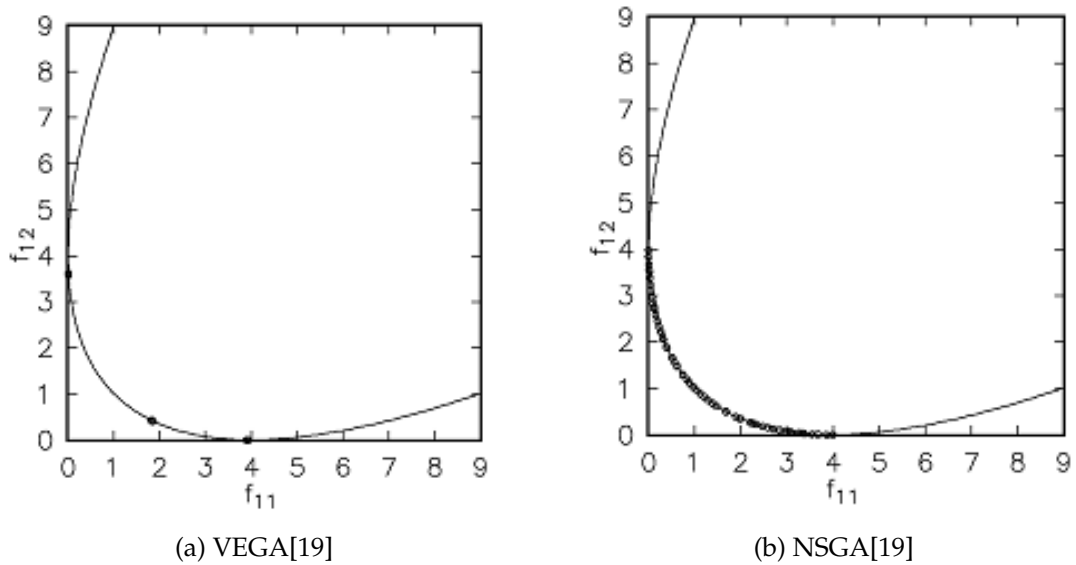


Figure 7.14: Solution of Shaffer's F1 (see eq. 7.17) after 500 generations using NSGA and VEGA, highlighting VEGA's tendency to fixate on particular solutions [19]

Algorithm 7.10 NSGA algorithm [19]

1. Generate an initial population \mathbf{P}
 2. For each generation $g \in \{0, 1, \dots, G\}$
 - (a) Evaluate using $\mathbf{F}(\mathbf{x})$
 - (b) Perform domination calculations to identify the non-dominated front f_1
 - (c) For each front f_n
 - i. Assign a large arbitrary fitness to members of f_n
 - ii. Calculate $\sum Share(d_{i,j})$ for each individual where $i =$ current individual and $j \in \{0, \dots, |\mathbf{P}|\}$ but $i \notin j$
 - iii. Perform sharing and reduce fitnesses f eq. 7.16 between all individuals
 - iv. Ignore f_n from now on
 - (d) Selection and Reproduction
-

NSGA is open to criticism from a number of directions: [20]

- Non-dominated sorting is computationally complex for large populations [20]

- Lack of elitism[20]
- The specification of sharing parameter $\sigma_{sharing}$ [20]

These issues are addressed with the refinement of NSGA - NSGAI.

7.6.6 NSGAI (Non-dominated Sorting Genetic Algorithm II)

NSGAI takes a simpler approach to specifying fitness from non-domination. Instead of assigning a dummy fitness and altering it due to crowding (as NSGA does), the fitness is set to the front index (beginning at 0). In a separate stage the crowding (in the solution space) is calculated using a density estimator, yielding a larger figure for more crowded individuals always < 1 . Adding this to the front index results in a final fitness incorporating domination and density information with the most dominated and crowded being least fit.

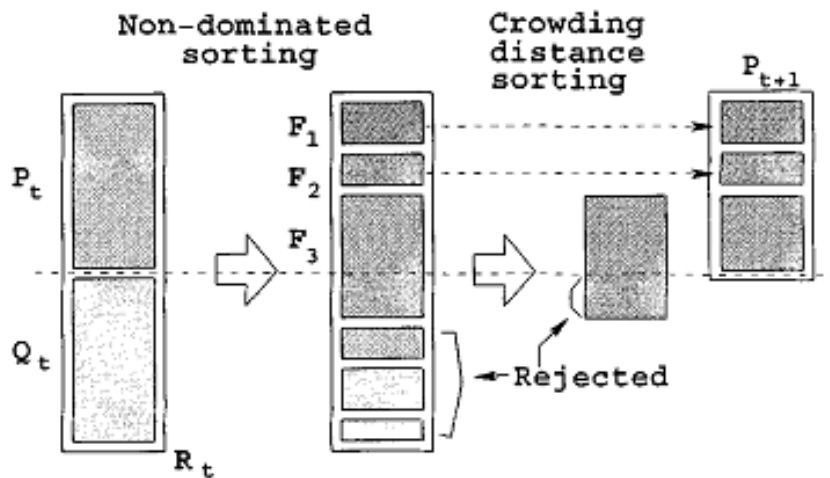


Figure 7.15: NSGAI procedure, where $P_t = P_p$ and $Q_t = P_\phi$ and $F_n = f_n$ [20]

Algorithm 7.11 NSGA II algorithm [20]

1. Generate an initial population \mathbf{P} of size $|\mathbf{P}_0|$
 2. For each generation $g \in \{0, 1, \dots, G\}$
 - (a) Evaluate using $\mathbf{F}(\mathbf{x})$
 - (b) Perform non-dominated sort on \mathbf{P} and assign increasing front numbers to \mathbf{f}
 - (c) For each front f_n until $\mathbf{P}_p \leq |\mathbf{P}_0|$
 - i. assign crowding distances of members f_n (increasing crowding \rightarrow lower fitness) to \mathbf{f} .
 - ii. $\mathbf{P}_p = f_n \cup \mathbf{P}_p$
 - (d) Sort \mathbf{P}_p using \mathbf{f}
 - (e) Select and Reproduce using \mathbf{P}_p to create and offspring population \mathbf{P}_ϕ
 - (f) $\mathbf{PP}^{(i+1)} = \mathbf{P}_\phi \cup \mathbf{P}^{(i)}$
 - (g) Truncate \mathbf{P} to $|\mathbf{P}_0|$ (see fig. 7.15)
-

NSGAI still experiences limitations when applied to differing problems, for instance Zitzler's T6 problem[21] (see sec. 8.5.2.2). The non uniformity of the test problem to some extent subverts NSGAI's dynamic and parameter-less crowding approach leaving well a defined but incorrect final solution. SPEA (see sec. 7.6.7) on the other hand solves this well but has a poorly defined front. Critically Deb et al. [20] note "that all ... approaches faced difficulties in solving this so-called highly epistatic problem".

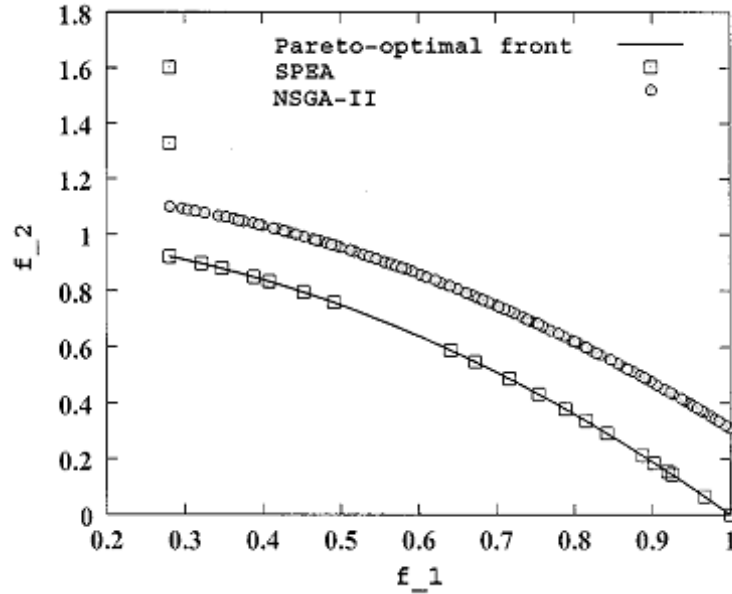


Figure 7.16: Comparison between NSGAI and SPEA in a minimisation problem where SPEA finds the solution better but NSGAI defines the front better on Zitzler T6[21] (see sec. 8.5.2.2)[20]

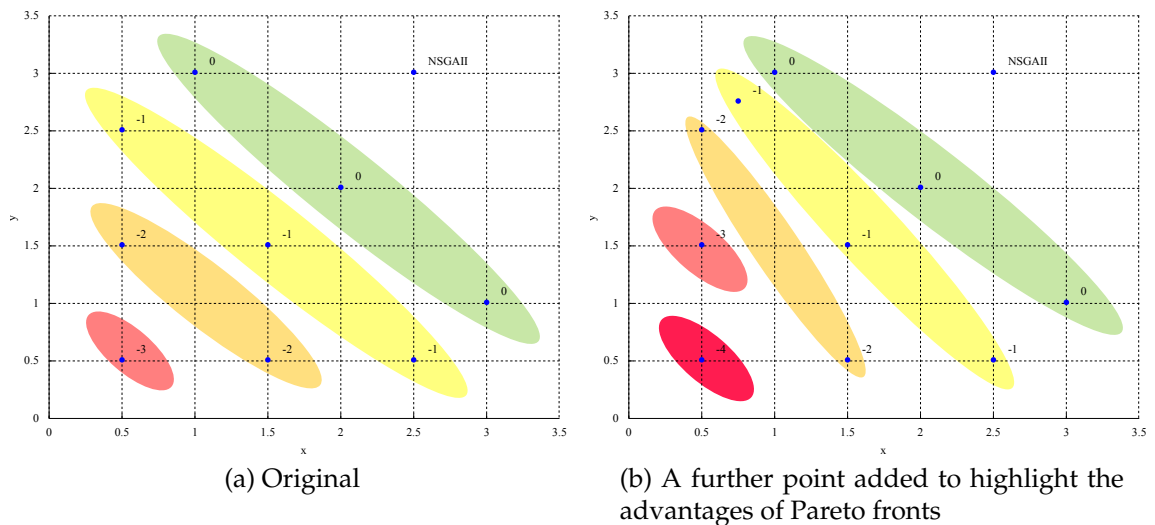


Figure 7.17: NSGAI’s fitness assignment (neglecting crowding). Colour groups are dominated by the same number of other solutions.

7.6.7 SPEA (Strength Pareto Evolutionary Algorithm)

The SPEA algorithm in contrast to NSGA alters the structure of the GA by storing the non-dominated solutions in an alternative population. It, too, uses the concept of domination to assign a fitness value. However instead of scoring

individuals on their Pareto front SPEA, determines the fitness from the number of members of the alternative (non-dominated) population who dominate them.

Members of both the population and alternative population participate in selection and so reproduction. Instead of detrimentally altering the fitness to promote a spread of solutions, SPEA identifies closely spaced solutions and removes them from the alternative population in a process called clustering.

Algorithm 7.12 SPEA algorithm [22]

1. Generate an initial population \mathbf{P} and empty alternative population \mathbf{P}_a
 2. For each generation $g \in \{0, 1, \dots, G\}$
 - (a) Evaluate using $\mathbf{F}(\mathbf{x})$
 - (b) Copy non-dominated members of \mathbf{P} to \mathbf{P}_a
 - (c) Remove newly dominated members of \mathbf{P}_a
 - (d) If $|\mathbf{P}_a| > a$ maximum then prune it using clustering
 - (e) Assign a fitness f to each individual in \mathbf{P}_a (see fig. 7.18)
$$f_{a(i)} = \frac{k}{|\mathbf{P}|+1}$$
 Where k for each individual ($\mathbf{X}_{a(i)} \in \mathbf{P}_a$) is the number of individuals $\mathbf{X}_j \in \mathbf{P}$ where $\mathbf{z}_i \succeq \mathbf{z}_j$.
 - (f) Assign a fitness f to each individual in \mathbf{P} (see fig. 7.18)
$$f_{(i)} = 1 + \sum_{j \in \mathbf{P}_a \cap \{\mathbf{z}_{a(j)} \succeq \mathbf{z}_{(i)}\}} f_{a(j)}$$
 meaning the fitness is the sum of all the fitnesses of the members of \mathbf{P}_a which dominate \mathbf{X}_j . One is added to ensure that f s of \mathbf{P}_a are lower than those of \mathbf{P} (f is to be minimised).
 - (g) Reduce the Pareto set using clustering
 - (h) Selection and Reproduction from $\mathbf{P} \cup \mathbf{P}_a$
-

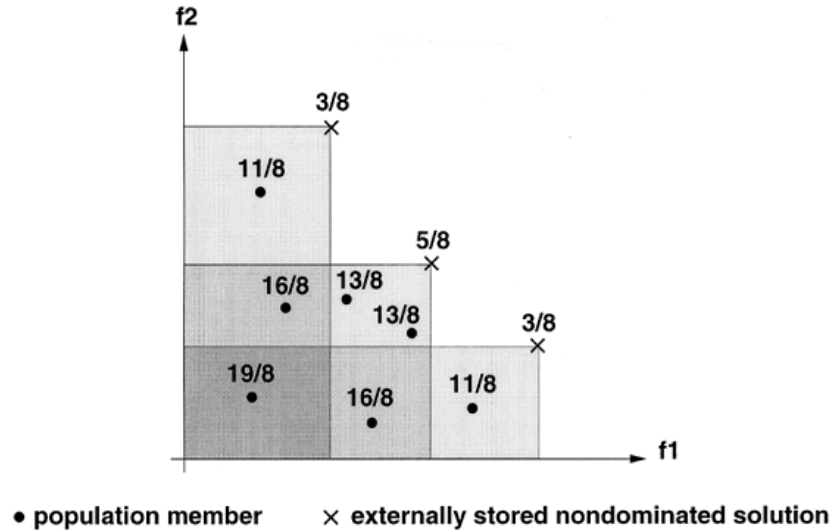


Figure 7.18: Fitness assignment in SPEA[22]

Once the true Pareto front of a problem has been found an algorithm based on SPEA is liable to find an infinite number of points on this front (if possible). SPEA attempts to avoid this problem by applying clustering to the Pareto front, which combines multiple points into a single point where the additional data has no value.[22]

Since its design SPEA has been criticised for certain aspects of its behaviour:[23]

- SPEA assigns identical fitness for individuals dominated by the same members of P_a . Where there is a single member of P_a the fitnesses of all members of P is the same.
- The clustering technique used can cause outer members of the Pareto front to be removed during truncation shortening the front.

7.6.8 SPEA2 (Strength Pareto Evolutionary Algorithm 2)

SPEA2 is a newer version of SPEA incorporating more recent refinements to overcome its drawbacks: [23]

- Improved fitness calculation based on both individuals dominating and dominated.

- Improved density estimation.
- Improved truncation to avoid losing boundary solutions.

Again an archive population \mathbf{P}_a is used, the size of this archive is fixed in contrast to the variable archive size as used in SPEA.

To avoid assigning individuals identical fitnesses as was possible in SPEA both the individuals dominated and those which dominate an individual are taken into account. Each individual is assigned a strength value related the number of solutions it dominates (see eq. 7.19):[23]

$$S(i) = \left| \left\{ \mathbf{z}(j) \mid \mathbf{z}(j) \in \{\mathbf{P} \cup \mathbf{P}_a\} \cap \left\{ \mathbf{z}(j) \succeq \mathbf{z}(i) \right\} \right\} \right| \quad (7.19)$$

where $|\cdot|$ denotes the cardinality of a set.

These are used to find the raw fitness of each individual, the sum of the strengths of the individuals which dominate it:[23]

$$R(i) = \sum_{j \in \{\mathbf{P} \cup \mathbf{P}_a\} \cap \{\mathbf{z}(j) \succeq \mathbf{z}(i)\}} S(j) \quad (7.20)$$

A fitness of 0 corresponds to a non-dominated individual and a high value corresponds to a highly dominated individual.

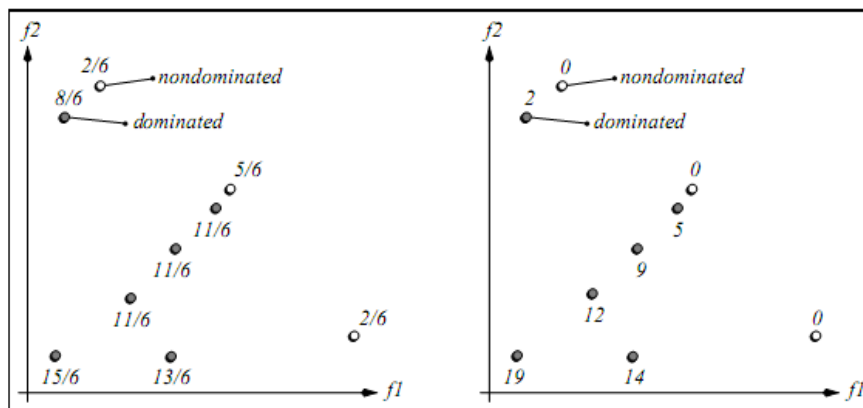


Figure 7.19: Comparison on the fitness assignment schemes of SPEA(left) and SPEA2(right). $R(i)$ is shown for SPEA2.[23]

Density estimation is taken care of using a slightly altered k^{th} nearest approximation which helps to preserve end points (see fig. 7.20). The numerical value found using this density estimation technique $D(i)$ is added to the raw fitness so the final fitness is:

$$\hat{f}_{(i)} = R(i) + D(i) \quad (7.21)$$

This solves the problem of truncation by reducing the fitness of densely packed individuals instead of clustering.[23]

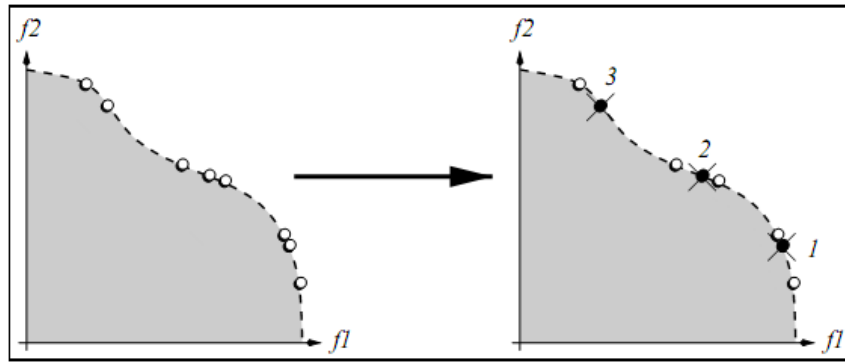


Figure 7.20: Demonstration of the truncation used by SPEA2 where the target population size is 5. k^{th} nearest crowding (see sec. 7.6.3) is used to rank the individuals which helps preserve end points.[23]

Algorithm 7.13 SPEA2 algorithm [23]

1. Generate an initial population \mathbf{P} of size $|\mathbf{P}_0|$ and an external population \mathbf{P}_a
 2. For each generation $g \in \{0, 1, \dots, G\}$
 - (a) Evaluate using $\mathbf{F}(\mathbf{x})$
 - (b) Assign a fitness to all members of $\mathbf{P}_{(g)}$ and $\mathbf{P}_{a(g)}$
 $\hat{f}_{(i)} = R(i) + D(i)$ where $R(i)$ is eq. (see eq. 7.20) and $D(i)$ is the clustering parameter.
 - (c) Copy all non-dominated members of \mathbf{P} to \mathbf{P}_a .
 If $|\mathbf{P}_a| > \max |\mathbf{P}_a|$ then truncate, sorting by fitness.
 If $|\mathbf{P}_a| < \max |\mathbf{P}_a|$ then add most fit dominated individuals.
 - (d) Selection and recombination from \mathbf{P}_a .
-

SPEA2 is better than SPEA in all test problems undertaken by Zitzler et. al.[23]. SPEA2 and NSGAI have the best performance over all[23]. In higher dimensional problems SPEA2 appears to have advantages over NSGAI[23].

Chapter 8

Evolutionary Algorithm

8.1 Klystron Optimisation

As in the manual optimisation the decision vector is the drift lengths and the frequencies of the cavities. To simplify the optimisation slightly the input cavity is fixed to the operating frequency and the input and output coupling are fixed to the specification values.

All of the drift lengths (5) and the remaining cavity frequencies (5) are free parameters (see table. 8.16). The solution vector will contain the efficiency, Δ_{gain} (see sec. 8.2.3), interaction structure length and the slowest electron exit velocity u_{slow} (see sec. 8.2.4) which are minimised or maximised as appropriate. This should yield an efficient, short klystron within the specification gain, with an exit velocity above the required minimum.

Table 8.1: Optimisation space definition

Vector	Components
Decision vector	[drift1, drift2, drift3, drift4, drift5, freq2, freq3, freq4, freq5, freq6]
Solution vector	[efficiency, length, Δ_{gain} , u_{slow}]

8.2 Klystron Simulation Issues

There are a number of issues that are unique or strongly related to the optimisation of klystrons.

- The multitude of short low efficiency klystrons (see sec. 8.2.1)
- Erroneous simulations (see sec. 8.2.2)
- Bandwidth optimisation (see sec. 8.2.3)
- Reflected electrons (see sec. 8.2.4)

Each of these requires an individual solution to either sidestep or solve the problem.

8.2.1 Pareto Search Pressure Focusing (PSPF)

Generic multi-objective EAs which use the Pareto front concept are often demonstrated in publications using mathematical equations. These often have not only a much clearer transformation between decision and solution space, but also and critically, every point on the first Pareto front is equally optimal. The interest in the full extent of the front is something which does not apply to klystron design.

Fig. 8.1 demonstrates the problem. Short (relatively) efficient klystrons are targeted as the algorithm been designed as such. However as a klystron becomes shorter it has a tendency to become less efficient, and there exists a significant set of decisions which result in a very short very inefficient klystron. Simply using the Pareto front concept these are *equally* as optimal as a longer significantly more efficient klystron.

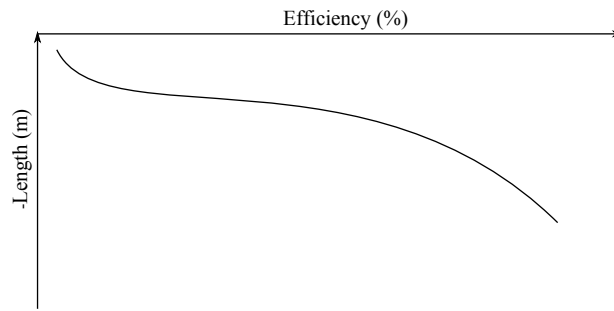


Figure 8.1: Relationship between length and efficiency

As short inefficient klystrons are easily discovered there is a clumping of solutions in this region. At best it could be hoped that there would still be an equal search pressure along the front. However where a clumping exists, purely due to weight of numbers, the chances of a selection from this region increases and the algorithm clumps further. Without modification the Pareto front concept does not usefully optimise a klystron interaction structure.

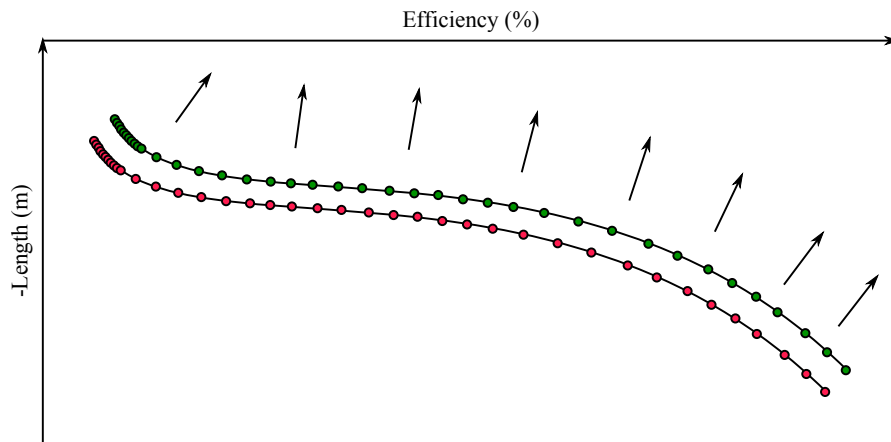
An adjustment to the Pareto concept, Pareto Search Pressure Focusing (PSPF), is proposed as selection is defined by Pareto front index and the Pareto concept still has useful behaviour.

The Pareto front's indexes are first separated by one to create empty fronts between them. An arbitrary cut-off point is defined as a fraction of the maximum in a single dimension (maintaining the separation between dimensions) above which the Pareto front index is maintained, but below which it is demoted to the front behind (by incrementing the front index). The focusing effect is felt though the entire range of fronts, so individuals above the cut-off are always preferentially selected.

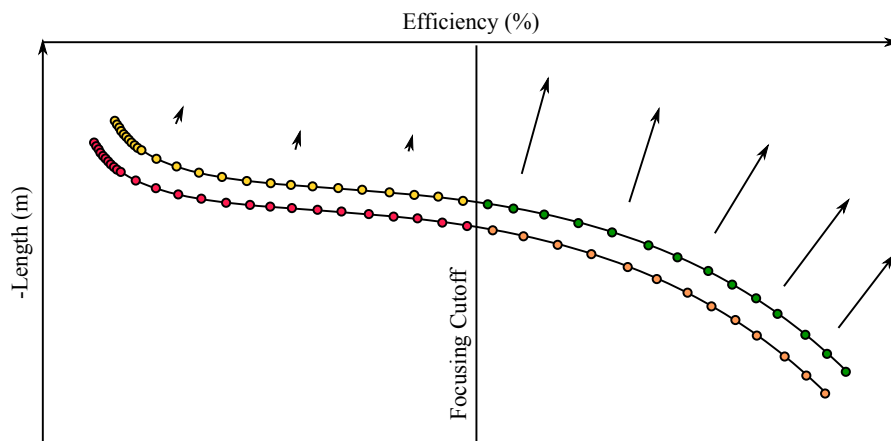
This is similar in target to solution space constraints discussed in [20], however it has the refinement of a variable cut-off, as solutions being found *above* the cut-off are, at least initially, the exception rather than the rule. A variable cut-off allows an increase in search pressure for the most efficient individuals from the start.

SPEA2 does not lend itself to such an adjustment, so its advantages in many

dimensions[23] have to be foregone. NSGAI's fitness assignment, being front based, is easily adapted to work with PSPF and is used in the rest of this work.



(a) Search pressure without PSPF



(b) Search pressure with PSPF

Figure 8.2: Demonstration of the effect of PSPF (Good→Bad) the arrows represent the search pressure (note the sign of the y axis)

8.2.2 Simulation Vetoes

Not all inputs to a klystron code translate to a valid output. There must be a mechanism to deal with decision vectors which:

- Use the maximum number of iterations in the calculation and so have not converged reliably
- Report an Efficiency lower than 0%

The algorithm uses a fairly simplistic solution to this problem by vetoing them. Without the Spartan concept it would be a simple matter of discarding the re-

sult. Instead the solution vector is set to be excessively poor in all objectives. This allows the Spartan operator to select them and use them to avoid the poor region of the decision space.

8.2.3 Bandwidth Targeting

Optimisation of the bandwidth presents two problems:

- Traditional computation of the bandwidth of each individual is computationally expensive if a small signal analytical solution is not used
- Optimisation *beyond* the specification bandwidth is unnecessary

Firstly the bandwidth can be defined in different terms. The precise bandwidth of each simulation is of no interest, but critically the algorithm must preferentially select individuals closer or equal to the specification bandwidth.

The gain is calculated in three positions; the centre and the upper and lower bandwidth points. The objective used for optimisation is in fact the sum of the drop between the centre and the bandwidth points in dB.

$$\Delta_{gain} = 10 \cdot \left[\log_{10} \left(\frac{lower}{center} \right) + \log_{10} \left(\frac{upper}{center} \right) \right] \quad (8.1)$$

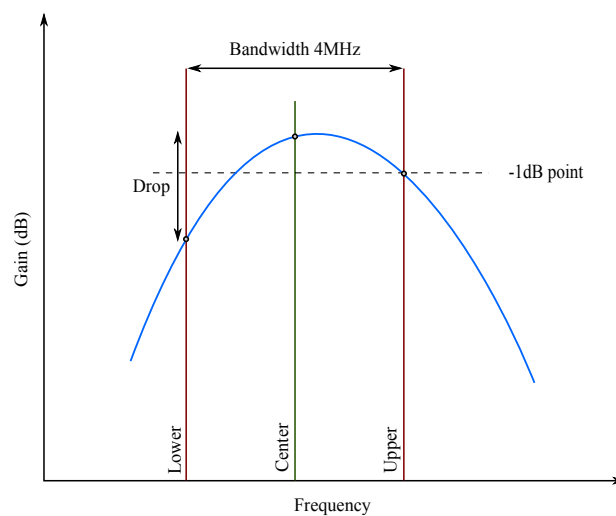


Figure 8.3: A more suitable definition of bandwidth for use in an evolutionary algorithm

Often scientific klystron bandwidths are defined at the -1dB point. As there is no value in optimising the bandwidth beyond the specification bandwidth it is capped at -1dB. All individuals within specification report a Δ_{gain} of -2dB (2×-1 dB) and as such are identically fit in this dimension, the intended behaviour.

Pareto search pressure focusing is used again to strongly target structures within specification; without this a full range of bandwidths could be targeted.

8.2.4 Reflected Electrons

As the klystron is optimised the output cavity will remove increasing amounts of power from the beam. This leads to increasingly slower electrons leaving the output gap. If the electrons leave the output gap too slowly there can be a build up of space charge which would effect the interaction. If more energy is removed from the beam some electrons could be reflected back up the tube back to the cathode bombarding its surface and shortening its life.

The veto approach could be used but often an individual which reflects electrons lies close to an individual which does not. As efficiency is related to the amount of power removed from the beam, targeting high efficiency klystrons will also tend to target klystrons with reflected electrons so a more elegant solution is required.

Instead the velocity of the slowest electron u_{slow} is defined as an objective. Using PSPF an absolute cutoff point can be defined below which solutions are deemed undesirable (see sec. 8.2.1). Thales use a figure of 8 kV, which equates to $5 \cdot 10^7 m/s$ [78].

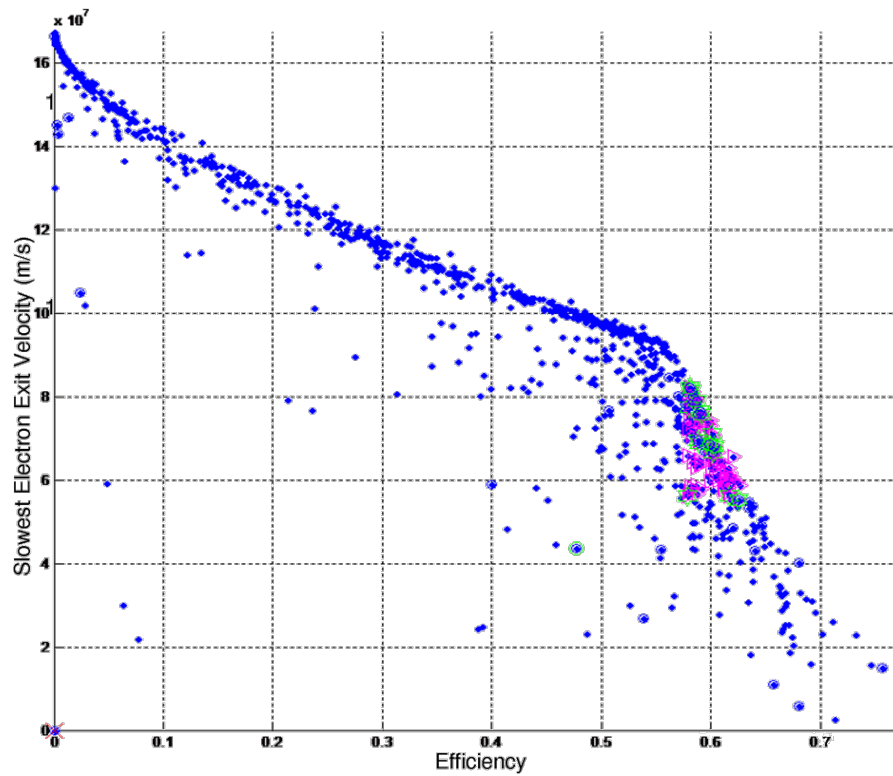


Figure 8.4: Optimisation of reflected electrons to avoid reflected electrons at the output gap (purple and green markings should be ignored). This is achieved by maintaining the exit velocity of the disks.

The plot displayed in fig. 8.4 contains the full length of the Pareto front. A knee is clearly visible at around 55% efficiency, above which efficiency drops rapidly for increasing u_{slow} . This is reduced using PSPF to the range on the efficiency axis above the knee as efficiency drops rapidly after this point.

8.3 Algorithm Refinements and Innovations

This section details the implementation of attempts to combine all of the current state of the art optimisation techniques with some new ideas. These new ideas take the form of improvements to the management of populations and efficiencies in the use of the computationally expensive simulation results.

8.3.1 Spartan Operator

In conventional evolutionary algorithms poor individuals are periodically discarded in order to improve the mean fitness of the population. They make no attempt to identify poor regions of the decision space using the information they contain. As considerable computational resources have already been expended to calculate the fitness of poor individuals it makes sense to try and make use of them in some way.[16]

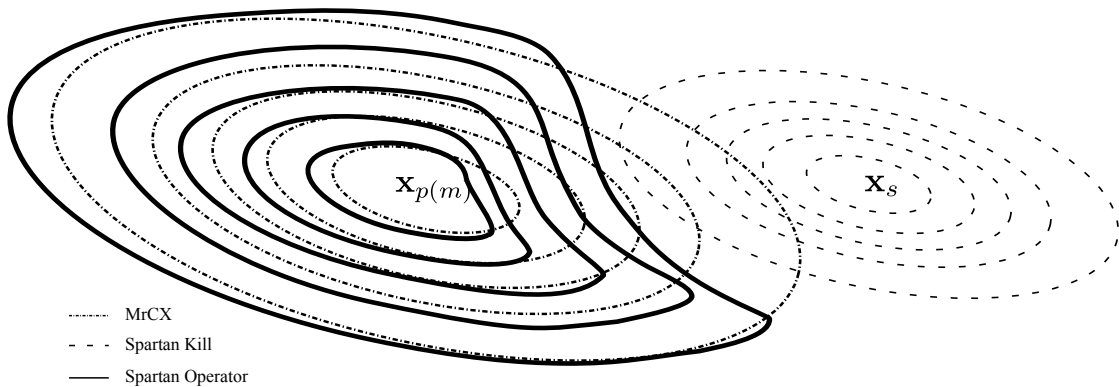


Figure 8.5: Probability distribution for MrCX alone compared with the distribution for the Spartan operator[16]

The Spartan operator attempts to perturb the probability distribution of the MrCX operator away from poor regions of decision space (see fig. 8.5). It does so by allowing poor members of the population to repeatedly kill offspring near them¹ until a surviving offspring is found. Members of the Spartan population \mathbf{P}_s are chosen to have a poor fitness (in contrast to the high fitness of the parent population \mathbf{P}_p).[16]

At the start of the Spartan procedure an offspring is generated using MrCX (see sec. 7.4.2) as before. Then every member of the Spartan population attempts to kill the offspring. It uses the same method as MrCX to create a random length vector (again scaled by the parent population) and multiplies this by the kill strength σ_s (as opposed to the search strength σ_x) in order to find the kill vector

¹In the ancient Greek city-state of Sparta the Elders would order the death of any infant judged too sickly or deformed to contribute to society. This reduced the considerable investment by the state to raise the citizen[85].

\mathbf{d}_k . This can be evaluated to discover whether the Spartan kills the offspring (see alg. 8.1 and fig. 8.6). This occurs when \mathbf{d}_k is larger than the distance between the offspring and the spartan. If the offspring survives this process then it should be well-spaced from the poor individuals in the population.[16]

The highest probability that an individual will be killed is if it appears at the exact location of a Spartan, decreasing as it moves away. If the offspring does not survive, a new offspring is created using the same parents but with a new set of Spartans. This is repeated until an offspring survives.[16]

The same parents (and mother) must be used to stop the search from being strongly repelled from a good location which is very close to a poor one, for instance a discontinuity. Instead as desired it perturbs the probability distribution away from the poor area rather than forbidding it completely.

Algorithm 8.1 Spartan algorithm[16] (see fig. 8.6)

For each member of \mathbf{P}_s :

1. Generate a random vector: \mathbf{d}'_k using MrCX algorithm (see alg. 7.6) based upon the parent population \mathbf{P}_p up to step 8.
 2. Multiply it by the kill strength: $\mathbf{d}_k = \mathbf{d}'_k \times \sigma_s$.
 3. Define a vector from the Spartan to the offspring $\mathbf{d}_s = \mathbf{x}_\phi - \mathbf{x}_s$
 4. Split \mathbf{d}_k into components perpendicular to and parallel to \mathbf{d}_s ($\mathbf{d}_{k\perp}$ and $\mathbf{d}_{k\parallel}$ respectively)
 5. If $|\mathbf{d}_{k\parallel}| \geq |\mathbf{d}_s|$, the offspring is killed (see fig. 8.6)
-

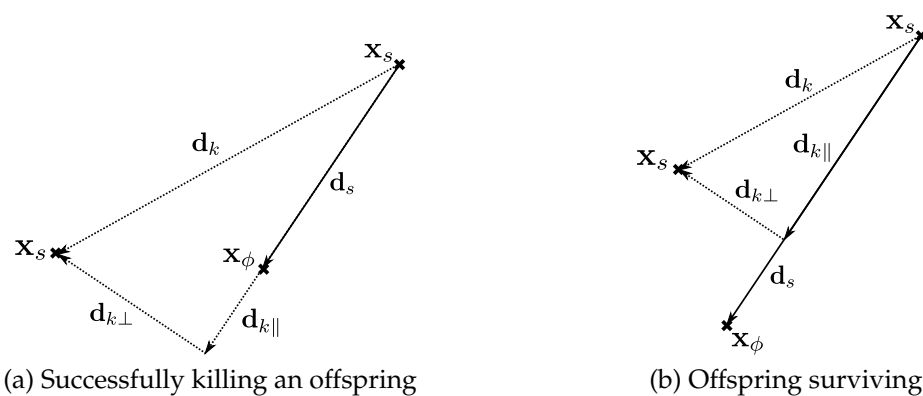


Figure 8.6: Examples of spartan behaviour involving two Spartans acting on two offspring in turn (a and b)[16]

8.3.2 Population Management

Population management has two functions, to limit the size of the population and to manage the insertion of offspring. The rationale for this is two fold:

- The population cannot be infinitely large
- Selectively removing poor members of the population to increase the mean fitness

The insertion method for single objective optimisation is often similar to evolution, in that each generation is replaced rather than added to. When individual evaluations are computationally expensive it makes sense to instead add to the population while limiting the size.

NSGAI (see sec. 7.6.6) uses a single population, the size of which is set arbitrarily at the start of the simulation and trimmed at the end of each generation. SPEA (see sec. 7.6.7) and SPEA2 (see sec. 7.6.8) complicate this by introducing a secondary population to contain the non-dominated solutions.

To avoid maintaining two populations the best solutions are not removed from the population, this concept is called elitism. For this reason the full set of non dominated solutions are designated to be elite. This has the disadvantage of over-populating the first non-dominated front once it has formed, however, as the number of evaluations is strongly limited but the computation intensity of the calculations and the number of dimensions is high, this is less of a problem.

An approach such as MGG (see alg. 7.7) or G3 (see alg. 7.8) is unsuitable in this case as more individuals are evaluated than are inserted. When the Spartan concept is used all information, good or bad, is important. In this work every individual evaluated is inserted.

8.3.2.1 Rejection

Work published by Deb [15] and Zitzler [86, 23] utilises rejection, a basic population management system. Simply that the population must not increase

beyond a preset level. On each insertion into the population another individual is removed, in this case the least fit (see sec. 8.3.2.1).

Algorithm 8.2 The rejection population management algorithm

input population $\mathbf{P}^{(g)}$ with fitness distribution f and max population size T

$$\mathbf{P}_i^{(g+1)} = \text{sort}_f \left(\mathbf{P}_i^{(g)} \right)$$

where $i \in \{0, \dots, T\}$

output $\mathbf{P}^{(g+1)}$

8.3.2.2 Significance

Rejection loses some of the advantage to be gained by the use of the Spartan operator, that of regions of poor fitness. The concept of significance is a measure of the number of times an individual is selected to reproduce or act as a Spartan. If an individual does not find itself selected often, relative to the rest of the population it then can be considered insignificant and removed.

For each generation the significance index (I_s) of each individual is incremented by one. The mechanism used to measure the relative selection frequency is the application of a multiplier of $0 \leq x \leq 1$ to the I_s . Thus a large I_s in fact denotes an *insignificant* individual, conversely a low I_s denotes a *significant* individual. An insignificant individual can exist after a number of generations because in previous generations there were less significant individuals

The significance paradigm can also be used in a subtly different way. Before the population is trimmed to the new level it is ordered using the “fitnificance”. This is a combination of the fitness (related to the Pareto front index) and the significance index. The effect of this is that the least significant members of the population are removed, and where the significance indexes match, the least fit member is removed.

If the population is controlled purely with fitness then the advantage of retaining poor individuals is lost. If instead of sorting the population using fitness,

significance is used and the least significant individuals rejected, this information can be maintained.

8.3.2.3 Significance Population Management

As rejection population management sets a arbitrary limit on the population, potentially important information can be lost. Significance population management limiting population size sets a limit on the significance index. This leads to a variable population size, theoretically tailored to the information represented.

Algorithm 8.3 The significance population management algorithm

input population $\mathbf{P}^{(g)}$ with fitness distribution f and max significance index I

$$\mathbf{P}_i^{(g+1)} = \text{sort}_\iota \left(\mathbf{P}_i^{(g)} \right)$$

$i \in \{ \text{index of individuals with significance } \iota < I \}$

output $\mathbf{P}^{(g+1)}$

8.4 Simulation Tools

8.4.1 High Throughput Computing Resources (CONDOR)

In order to benchmark MOEAs² it is usual to use 20-25,000 calculations to ensure good convergence.[20] Although the convergence need not be as thorough for this problem a large number of simulations must be undertaken. As shown in sec. 8.4.2 each simulation with a suitable degree of accuracy will take approximately 5 minutes. This taken with a suitable number of iterations and the conclusion is simple:

$$5 \text{ minutes} \cdot 10,000 \approx 35 \text{ days} \tag{8.2}$$

²Multi-Objective Evolutionary Algorithms

This is simply not a plausible proposal for a single desktop processor; more powerful resources are needed.

Although high power resources can perform large complex tasks quickly, due to their configuration they are poorly suited to efficiently completing many low intensity problems. Due to their expense and utility for large problems, processor time on such a system is at a premium.

The antithesis of such a system is a *high throughput* computing resource. This is a system tuned to rapidly performing a large number of small, largely isolated tasks. Their load managing systems also help to maintain availability of the system to other users.

At the Cockcroft Institute there is a small grid system using idle time from the Institute's desktop PCs. At any one time there is access to 30-50 Windows cores. Condor high throughput computing management system is used to control the grid.[87]

The easy access to these cores reduces the optimisation time to a reasonable level:

$$\frac{5 \text{ minutes} \cdot 10,000}{50} = 16 \text{ hours} \quad (8.3)$$

MatLab has interfaces for several grid computing architectures, but unfortunately Condor is not among them. However there is an open source alternative interface written in Java by Hidemoto Nakada.[88] Condor Java API provides an interface to allow Java to submit jobs to the Condor pool and monitor their status in the pool. A failure status allows the job to be resubmitted, and a success status to return the data to the MOEA for analysis.

8.4.1.1 Asynchronous Calculation

Use of the Condor pool and the inconsistent nature of the length of each calculation points us to the unsuitability of the generation concept. If the standard generation concept is used and a set of decision vectors submitted to the

Condor pool they *all* must complete before handing the results back to the algorithm. This markedly lowers the computational efficiency of the optimisation. If a single calculation is performed per generation then the performance advantage of grid computing is lost.

Instead an asynchronous calculation model is used (see fig. 8.7). As in a generational model a set of individuals are submitted for calculation, but when the first is evaluated this is handed back to the EA. The EA returns a single new individual vector to replace it. In this implementation multiple solution vectors can be accepted back by the EA, and multiple decision vectors returned.

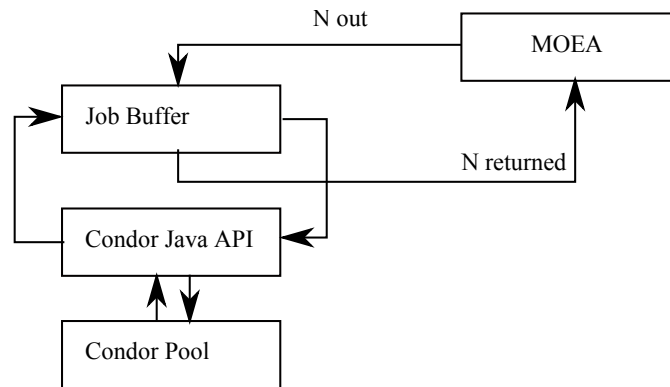


Figure 8.7: Schematic representation of asynchronous calculation

A useful advantage of this, quite apart from computational efficiency, is that the information in the calculation buffer is always as up to date as possible. This is similar to the difference between the Gauss and Gauss-Seidel algorithms for solving simultaneous equations. This is an interesting example of the underlying computing architecture informing and actually improving the design.

8.4.2 AJDisk

The algorithm of course, being generic, does not have klystron simulation ability so must be paired with a code capable of this.

Many large signal models are only available to a restricted range of users for reasons of commercial or national confidentiality. The best code available

for this work was Klys2D but that was only accessible at the Thales factory in Velizy. Klys2D has been validated by comparison with experimental results. However a simpler code, AJDisk³ is available without license and is used as the principal tool used for interaction modelling in this project.

AJDisk lends its self to use in an EA as it is easily distributable, license free, its run-times are measured in 100s of seconds and it is configured by plain text files.

8.4.2.1 Disk/Ring Model

AJDisk is a disk/ring based single beam klystron simulation code.

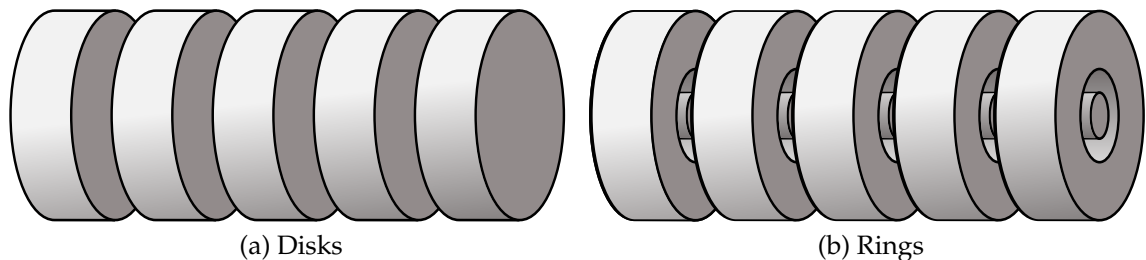


Figure 8.8: The two space charge representations used by AJDisk to model the beam

AJDisk reduces the klystron simulation problem into three sections.[24]

- Equations of motion
- Cavity fields
- Space charge

8.4.2.2 Equations of Motion

AJDisk uses a Finite Difference Time Domain method to solve the equations of motion, that is to say that the velocity and position are updated for each time-step with reference to the previous time-step.[24]

³Kindly provided by Aaron Jensen, SLAC

The update equations are derived by combining the equation of motion eq. 8.4 with the Lorentz force eq. 8.5.

Equation of motion:[24]

$$\mathbf{F} = \frac{d(\gamma m \mathbf{v})}{dt} \quad (8.4)$$

where m = mass, t = time, \mathbf{v} = velocity, γ = Lorentz factor

Lorentz force:[24]

$$\mathbf{F} = q (\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (8.5)$$

where, q = charge, \mathbf{E} = electric field \mathbf{B} = magnetic field

Which yields update equations for position (see eq. 8.6) and velocity (see eq. 8.7).[24]

$$z^{n+1} = \frac{1}{\gamma^3} \frac{q}{m} (\Delta t)^2 E_z + 2z^n - z^{n-1} \quad (8.6)$$

$$v^{n+1} = 2 \frac{1}{\gamma^3} \frac{q}{m} (\Delta t) E_z + v^{n-1} \quad (8.7)$$

8.4.2.3 Cavity Fields

It is the fields in the cavity which modulate the beam by accelerating and decelerating the beam.

These fields can be approximated using a Gaussian profile (see eq. 8.9) defined by the shaping factor k , which is related to the longitudinal gap coupling factor M averaged over the beam radius (see eq. 8.9). This field is assumed to be constant radially, which is not the case in reality.[24]

$$E_{cav} = |V| f(z) \cos(\omega_0 t + \theta) \quad (8.8)$$

$$f(z) = \frac{k}{\sqrt{\pi}} \exp\left[-k^2 (z - z_0)^2\right] \quad (8.9)$$

$$k = \frac{1}{2} \frac{\beta_e}{\sqrt{-\ln(M)}} \quad (8.10)$$

[24] states that the phase and amplitude can be found using only the first harmonic. A Fourier transform (w.r.t. time) is performed on the total current. where Q is the charge on a disk and v is the disks velocity.

$$I_{ind,1} = \mathcal{F}\{I_{ind}(t)\} = \mathcal{F}\left\{\sum^i Q_i v_i(t) f(z_i)\right\} \quad (8.11)$$

As the phase and amplitude are equivalent to the electric field, Ohms law and circuit theory can be used to find the amplitude and phase. [24]

$$V = I_{ind,1} Z_{cav} \quad (8.12)$$

Where the cavity impedance is:[24]

$$Z_{cav} = \frac{R/Q}{\frac{1}{Q_i} + j \frac{\omega_0^2 - \omega_{cav}^2}{\omega_0 \omega_{cav}}} \quad (8.13)$$

A solution is found by running a series of disks through the cavity using the above equations and iterating until the values for phase and amplitude of the cavity fields converge to solutions [24].

8.4.2.4 Space Charge

The space charge calculation is based on reducing the disks to delta function disks (perfectly thin). This does not effect the simulation when the disks are spaced but when disks overtake it leads to a singularity in the force. The

electric field distribution for such a disk in a drift tube can be calculated in three steps:[24]

- solve Laplace's equation in cylindrical coordinates
- find a Green's function to represent the potential due to a unit charge in an infinitely long conducting cylinder
- integrate the Green's function over the "delta-function disk" to find the electric field due to all the electrons in a single disk

This calculation yields the z component of the electric field at a position z due to a disk of charge Q at position z_0 with radius b in a drift tube of radius a which allows the force between the disks to be calculated.[24]

$$E_z = \frac{2}{\pi b} \frac{Q}{2\epsilon a} \sum_{l=1}^{\infty} \exp \left[-\frac{\mu_l |z - z_0|}{a} \right] \frac{J_1 \left(\mu_l \frac{b}{a} \right) J_0 \left(\mu_l \frac{r}{a} \right)}{\mu_l J_1^2(\mu_l)} \text{sgn}(z - z_0) \quad (8.14)$$

where μ_l is a constant.

8.4.2.5 Calculation Structure

These results are combined in AJDisk in the following manner (see fig. 8.9).

n = Cavity Number
 T = One RF Period
 $E_{cav_n} = |V_n| f_n(z) \cos(\omega t + \theta)$
 $f_n(z)$ = Field Profile of Cavity n
 i = Disk Number

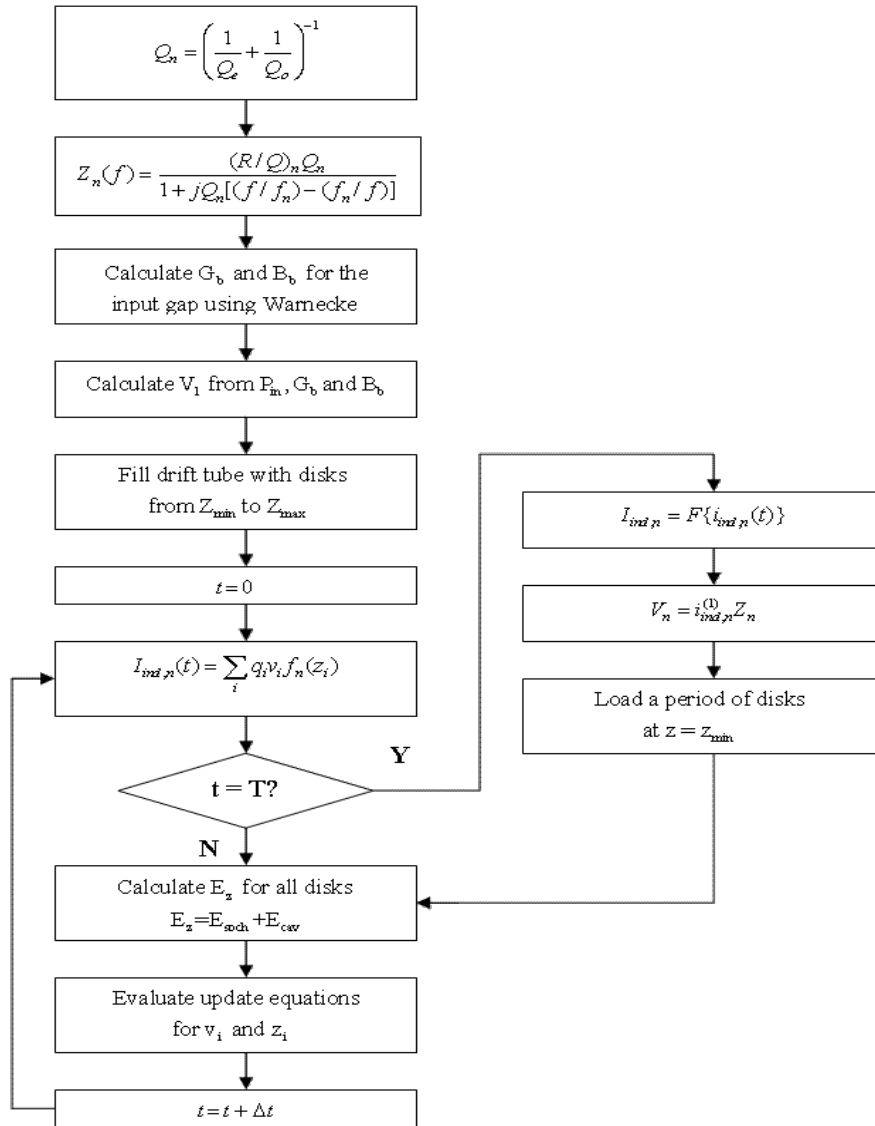


Figure 8.9: Flow chart showing AJDisk’s operation, reproduced from [24]

8.4.2.6 Convergence

The method of solution used by AJDisk requires that the number of disks and rings are chosen. The results of a convergence study on these two parame-

ters are shown in fig. 8.10 and 8.11.

The first consideration is the convergence of the output power. This is appropriate as a figure of merit as it is also the equivalent of gain and efficiency (for a fixed klystron configuration).

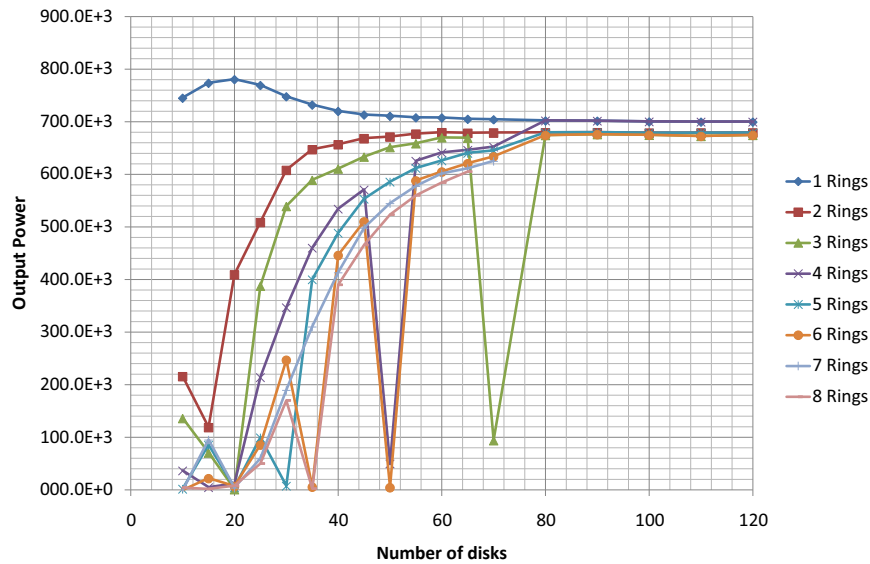


Figure 8.10: Number of disks and rings against reported output power

The sudden drops in output power experienced in fig. 8.10 are unexplained, but critically only occur for numbers of rings over 2; for this reason only 1 or 2 rings are considered.

It should be noted that the simulations converge to one of two final values, and as separate numbers of rings converge to different values, it is difficult to select one value over the other. Of the number of rings that converge, two converge to the higher value and four to the lower. However as 1 and 4 rings converge to the same value it is hard argue for either value. As a result the final results must viewed with an uncertainty of at least 3%. Note that 7 and 8 rings fail to converge for high numbers of disks using these settings. This can clearly be seen in fig. 8.11.

The introduction of rings to the simulation does not in fact allow AJDisk to model the interaction of the beam with the cavities more accurately. As there is no radial field variation each ring (if arriving in the same phase) will see the

same field. Its only impact is in the modelling of the space charge in the drift tubes.

Another important consideration, due to the large number of simulations required, is the calculation time. As the Condor pool on which this convergence study was performed is made up of dissimilar machines, it is difficult to compare raw run-times. For this reason the run-time of the simulations shown is normalised using a benchmark calculation of known run-time. When plotted on a logarithmic scale a clear correlation can be seen. For a fixed number of disks increasing the number of rings increases the calculation time. Similarly for a fixed number of rings increasing the number of disks increases the calculation time.

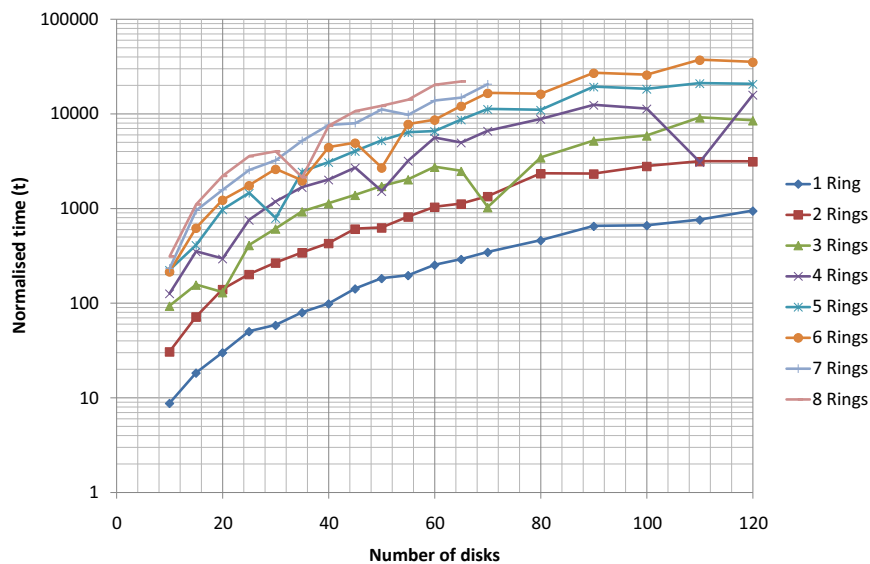


Figure 8.11: Number of disks against *normalised* time. The time units used are comparable to seconds for a modest CPU.

If 2 rings were chosen then the calculation time is increased by approximately a factor of 4 (see fig. 8.11). For this reason 1 ring is used (a disk) in conjunction with 80 disks.

8.5 Methodology and Calibration

This EA, in common with most other evolutionary algorithms, has a large number of parameters, all of which must be chosen. As far as possible these are evaluated for a range of values and the optimum chosen.

Each setting is attributed an arbitrary default value until it has been evaluated, when the optimum value is taken. For simplicity parameters not directly connected to multi-objective optimisation are optimised in single objective.

8.5.1 SpartanEA Implementation

There are a wide number of possible configurations of an evolutionary algorithm. As stated in sec. 8.2.1 NSGAI with PSPF is used to assign fitness to the population. For this implementation tournament selection has been chosen for its simplicity and beneficial characteristics (see sec. 7.3.2). MrCX has been chosen over PCX for its improved handling of highly dimensional problems and the Spartan operator included.

The SpartanEA has been implemented in JAVA to accommodate the new concepts and easily communicate with the Condor pool (see fig. 8.12).

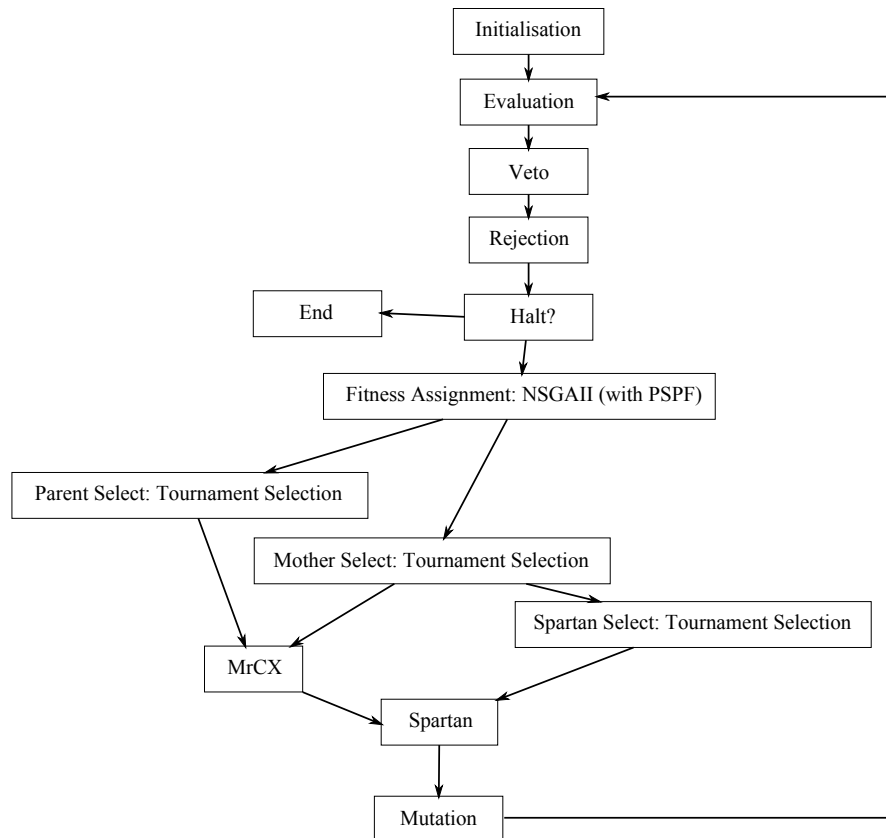


Figure 8.12: A flow chart representing the flow of populations around the SpartanEA

8.5.2 Test Problems

Calibrating the EA using klystron simulations would be a time consuming process. Instead it has been initially calibrated using a number of standard test problems. The first being a simple problem with N input values and a single output; the second being similar in behaviour to a klystron simulation.

8.5.2.1 Ellipsoidal Problem

The Ellipsoidal problem is a simple mono-modal problem that can be extended to any number of dimensions with the solution at $\mathbf{x}_{(i)} = 0$. Although it has no particular similarity to klystron simulations, it has been chosen for its simplicity.

It takes the form of the sum of a quadratic in each dimension, multiplied by

a constant (see eq. 8.15). By considering the 2 dimensional version all contours of the solution will take the form of an ellipse.[18]

$$F(x) = \sum_{i=1}^N i \cdot x_i^2 \quad (8.15)$$

Table 8.2: Elliptical problem properties

Parameter	Symbol	Value
Number of Dimensions	N	20
Initialisation range		[-10 : -5]
Search Boundary		$[-\infty : \infty]$

8.5.2.2 Zitzler T_6

This test function is described as including:

“two difficulties caused by the non-uniformity of the search space: first, the Pareto-optimal solutions are non-uniformly distributed along the global Pareto front (the front is biased for solutions for which $f_1(x)$ is near one); second, the density of the solutions is lowest near the Pareto-optimal front and highest away from the front.”[21]

$$f_1(x) = 1 - \exp(-4x_i) \sin^6(6\pi x_1) \quad (8.16)$$

$$g(x_2, \dots, x_n) = 1 + 9 \cdot \left(\left(\sum_{i=2}^m x_i \right) / (m-1) \right) \quad (8.17)$$

$$f_2(f_1, g) = 1 - \left(\frac{f_1}{g} \right)^2 \quad (8.18)$$

where $f_1(x)$ and $f_2(f_1, g)$ are the two objectives of the problem.

This problem is well suited to act as a proxy for a klystron simulation due to the non-uniformly distributed solution along the first Pareto front and the density of the solutions decreasing towards the Pareto front. The number of input dimensions is set to a number expected to be common for klystron optimisation and the number of objectives kept at 2 for simplicity.

Table 8.3: Zitzler T_6 problem properties

Parameter	Symbol	Value
Number of Dimensions	N	10
Initialisation range		$[-0 : 1]$
Search Boundary		$[-0 : 1]$
Objectives	M	2

8.5.3 Performance Metrics

When evaluating the performance of a single objective algorithm the performance can be measured trivially by inspecting the fittest and least fit individuals along with the mean. Evaluation of a multi-objective algorithm is a more complex matter.

It is simplified markedly by full knowledge of the solution. Deb et al. [20] for instance find the average minimum distance to a set of generated equally spaced set of points on the solution. This of course is of little value for a practical application, if the solution was well known the algorithm would not be needed. Although the generic case is relatively uninteresting, the klystron codes are sufficiently complex to invalidate this metric.

8.5.3.1 Domination Metric

A proposed generic performance metric is to evaluate the dominated area underneath the first Pareto front.

The rectangular rule of integration is traditionally inferior to, for instance, the trapezoid rule which better approximates the curve applied. This approximation of an unknown curve as shown in fig. 8.13 gives the impression that the algorithm has performed better than it in fact has. If an additional point (in red) is added to fig. 8.13 this metric would show no improvement when the front is in fact better defined.

Rectangular integration closely follows the \preceq operator used to calculate dominance (see fig. 7.13). The area reported by rectangular integration is the area

dominated by the points making up the first Pareto front when, and only when, the top-right corner approximation is used (see fig. 8.13b). The introduction of an additional point (shown in red) increases the value of this metric as desired (see fig. 8.13b).

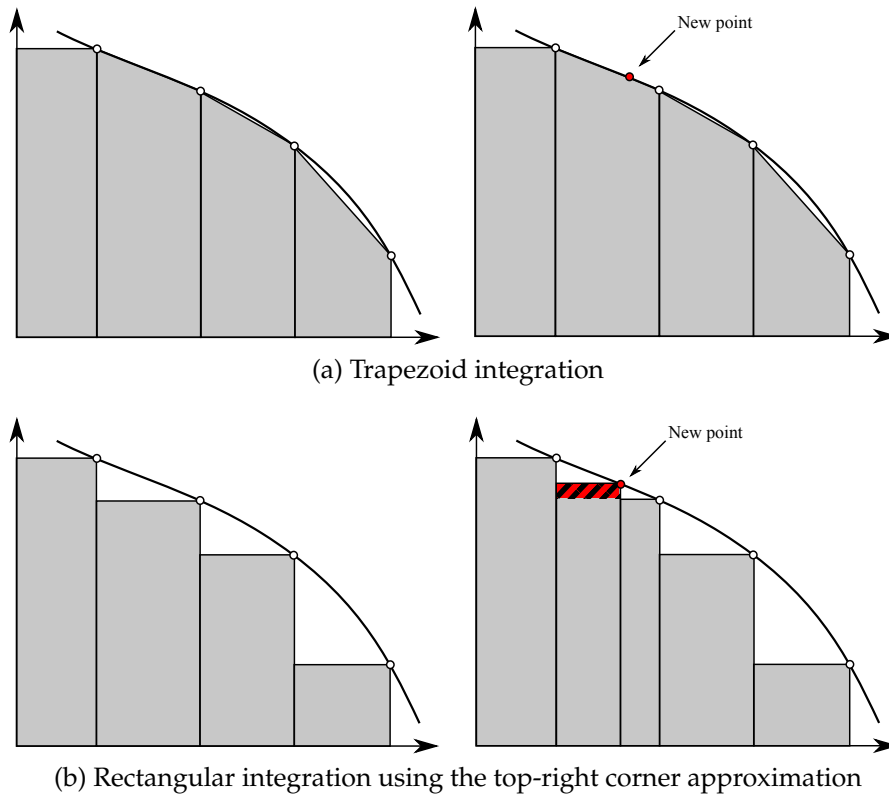


Figure 8.13: Methods of evaluating the area dominated by a number of points

The method proposed is not fully generic as for simplicity it applies to only two dimensions. It could be extended to N -dimensions by calculating a volume or a hyper-volume.

8.5.4 EA Setup

To attempt to isolate any parameter changes from more complex effects added by refinements to the algorithm the simplest possible configuration is used. Where there are no Spartans; mutation is disabled (0 probability) and no PSPF is implemented.

The initial optimisation parameters (see table. 8.4) either have simplification

in mind (the lack of Spartans and mutation) or have been instructed by initial work using single objective optimisation.[16]

Table 8.4: Initial calibration parameters

EA Parameter	Symbol	Value
Number of Spartans	s	0
Population management method		Reject (fitness)
Population Size	$ \mathbf{P}^{(0)} $	200
Mother Selection	$[\omega]_m$	Best in population
Parent Selection	$[\omega]_p$	Binary Tournament
Number of parents	p	11
Search strength	σ_x	0.7
Mutation Strength	σ_μ	0
Mutation Probability	$P(\mu)$	0
Number of generations	G	25,000
Number of Spartans	s	0
Spartan kill strength	σ_s	0

8.5.5 Result Presentation

Due to the probabilistic nature of the EA a single run of each parameter in a sweep is insufficient as the results can vary greatly. Instead for each value a total of 100 runs of the EA are performed unless otherwise stated.

To present the results a box plot (see fig. 8.14) is used with a plot of the mean superimposed. The median is shown as a red line with the upper and lower quartiles bounded by a blue box. The full range is shown by the black bars connected with dotted lines, while outliers (beyond the 1.5 quartile) are shown as red crosses.

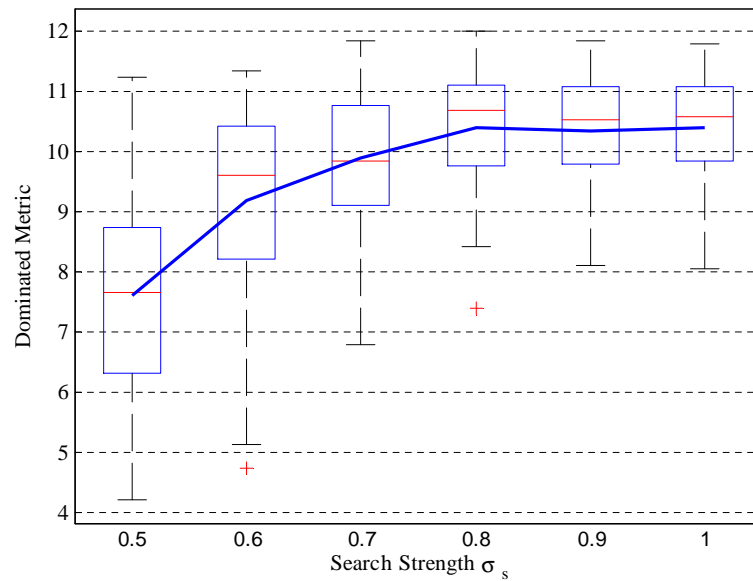


Figure 8.14: An example of a box plot (or box-and-whisker diagram) with an overlaid plot of the mean

8.5.6 Search Strength

The search strength defines how large the probability distribution is around the selected parents. If this value is too high then the offspring will be excessively spaced from the parents. This could cause the algorithm to fail to converge. For a low value the offspring may be too close to the parents, preventing the algorithm from evaluating new space.

Table 8.5: Single objective sweep of σ_x simulation setup [16]

EA Parameter	Symbol	Value
Search Strength	σ_x	{0.5, 0.6 . . . 1.5}
Test problem	$\mathbf{F}(x)$	Ellipsoidal
Number of generations run for	G	10,000
Population Size	$\mathbf{P}^{(0)}$	100

Using existing information of a single objective optimisation of the ellipsoidal problem the range of possible search strengths can be narrowed. As can be seen from fig. 8.15, a value of σ_x between 0.7 and 0.8 provides optimum

convergence efficiency. As this is a single objective mono-modal problem, all simulations were able to converge to the optimum.

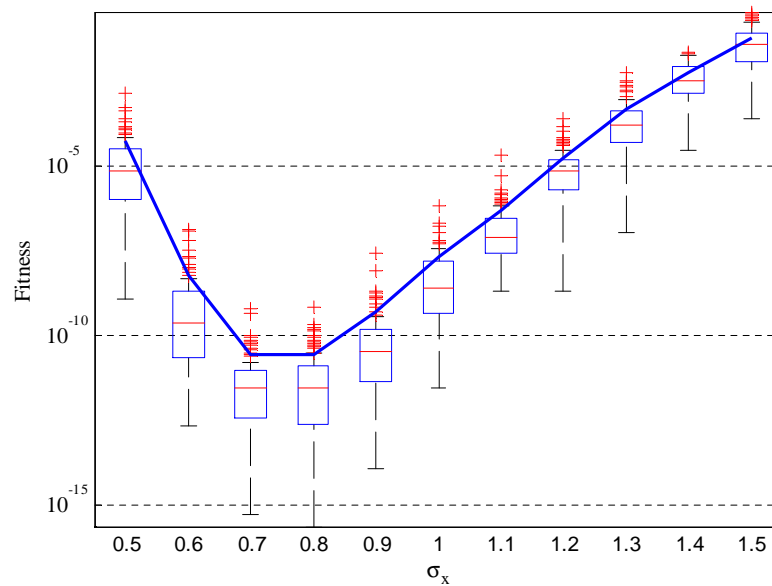


Figure 8.15: Single objective search strength σ_x sweep performed on the ellipsoidal problem[16]

Instructed by the single objective results a smaller range can be taken for the more time consuming multi-objective problem using the parameters described below:

Table 8.6: Search strength σ_x calibration setup

EA Parameter	Symbol	Value
Search Strength	σ_x	$\{0.5, 0.6, \dots, 1\}$
Test problem	$F(x)$	T_6

Problem T_6 was run over a range $\{0.5, 0.6, \dots, 1\}$, 100 times at each point to find any persistent trends in the domination metric. Clearly the algorithm is more successful in optimising this problem for search strengths over 0.8, becoming slightly less consistent for higher values (see fig. 8.16).

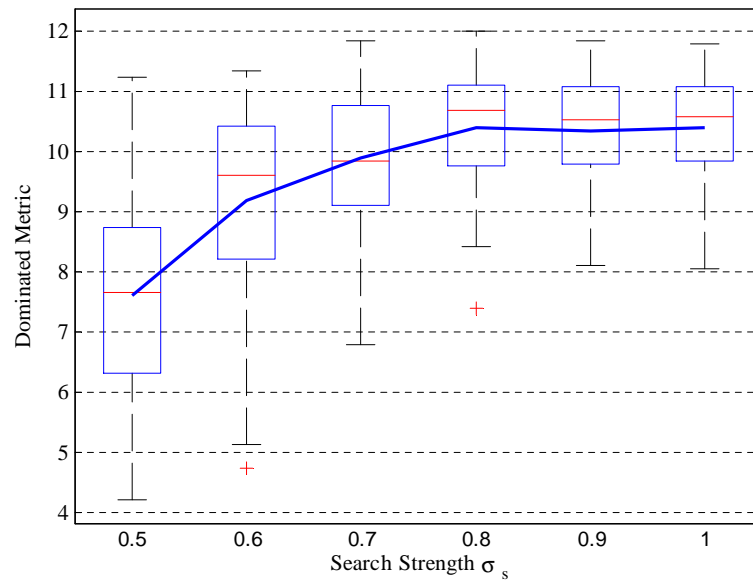


Figure 8.16: Multi-objective search strength σ_x sweep performed on T_6 . Showing that a search strength of 0.8 improves the dominated metric.

Table 8.7: Updated calibration parameters

EA Parameter	Symbol	Value
Search strength	σ_x	0.8

8.5.7 Mutation

The T_6 problem is a taxing problem where the algorithm can become stuck in multi-modal areas before the full solution. If mutation does not help the algorithm converge in this problem then it is unlikely that it will be advantageous for klystron simulations.

If the search strength is calibrated correctly there should be no need for mutation as MrCX itself contains a random element.

Table 8.8: Mutation strength σ_μ calibration setup

EA Parameter	Symbol	Value
Mutation Strength	σ_μ	$\{0, 0.3, \dots, 3\}$
Mutation Probability	$P(\mu)$	0.1
Test problem	$\mathbf{F}(x)$	T_6

The mutation operator does not appear to appreciably alter the dominated metric for the problem T_6 (see fig. 8.17). At most, the results can be considered inconclusive. With this in mind the mutation strength and probability are set to 0 (see table. 8.9) and the choice of search strength is justified.

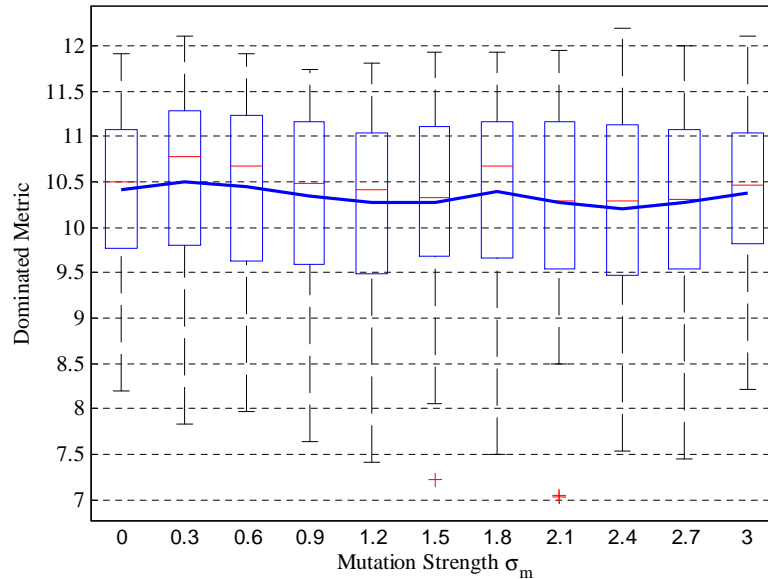


Figure 8.17: Multi-objective mutation strength σ_μ sweep performed on T_6 . Showing that mutation has at best an inconclusive effect on the dominated metric.

Table 8.9: Updated calibration parameters

EA Parameter	Symbol	Value
Mutation Strength	σ_μ	0
Mutation Probability	$P(\mu)$	0

8.5.8 Number of Parents

The number of parents per offspring affects the way in which the reproduction operator functions. With too few it can lead to an incomplete picture of the search space. The use of orthonormal vectors generated from the parents in PCX and MRCX would suggest that $N + 1$ parents would be necessary (see sec. 7.4.1).[15]

Table 8.10: Number of parents p calibration setup

EA Parameter	Symbol	Value
Number of parents	p	$\{1, 3, \dots, 29\}$
Test problem	$F(x)$	T_6

With fewer parents than 7 the algorithm struggles to converge well; with 7 to 17 the maximum convergence is almost indistinguishable; with the mean falling away as the number of parents increases (see fig. 8.18).

The $N + 1$ rule falls between these bounding points and for the purposes of this calibration will be taken as validation. With a variable dimensionality of the klystron problem such a rule is preferable to a single value.

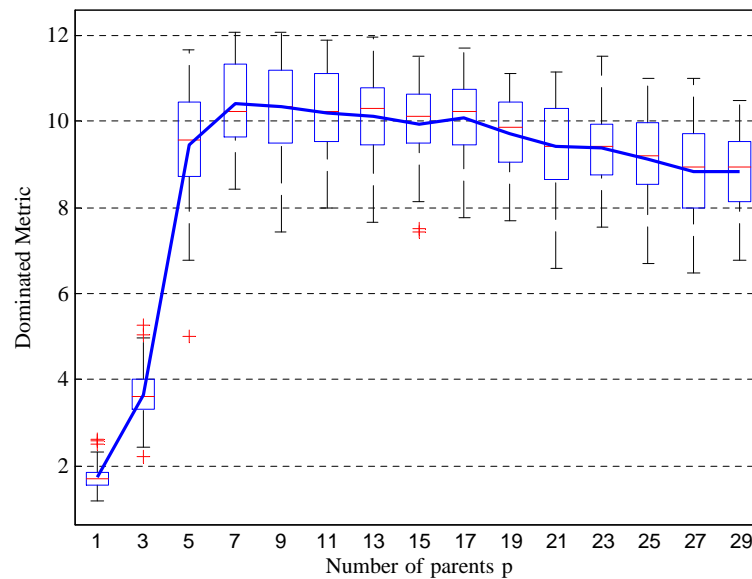


Figure 8.18: Multi-objective number of parents p sweep performed on T_6 . Showing that the original assumption of $N + 1$ holds.

Table 8.11: Updated calibration parameters

EA Parameter	Symbol	Value
Number of parents	p	11 (or $N + 1$)

8.5.9 Population Calibration

If a population is too small then it will lack diversity, and because of this will struggle to converge reliably. If the population is too large the selection operators struggle to operate effectively and at best there is no advantage.

Table 8.12: Population calibration setup

EA Parameter	Symbol	Value
Population Size	$\mathbf{P}^{(0)}$	$\{50, 100 \dots 600\}$
Test problem	$f(x)$	T_6

Problem T_6 was run over a range $\{50, 100 \dots 600\}$ 100 times at each point to find any persistent trends in the domination metric. Clearly the algorithm is more successful in optimising this problem for populations of 200 and above, becoming more consistent, but less successful for higher values.

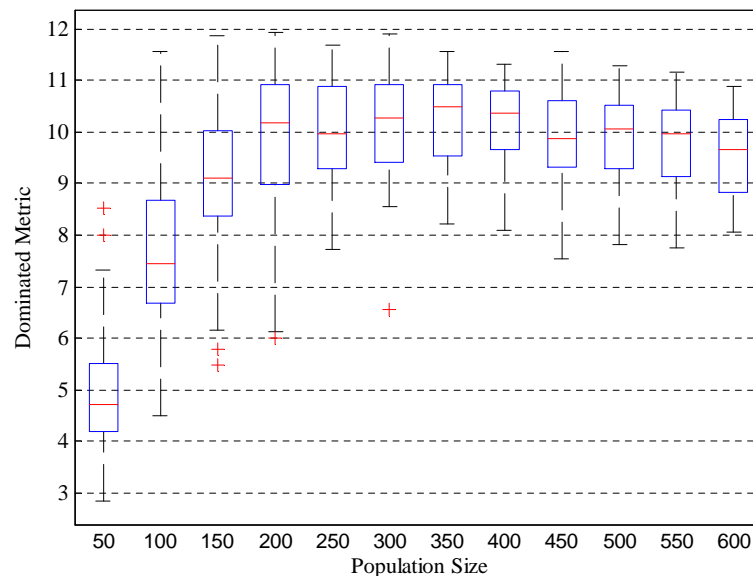


Figure 8.19: Multi-objective initial population size $\left| \mathbf{P}^{(0)} \right|$ sweep performed on T_6 . Showing that the optimal initial population is 300.

The optimal population size from this analysis is 300 as it not only reaches the highest value of the domination metric but does so more consistently as evidenced by the smaller range (see fig. 8.19).

Table 8.13: Updated calibration parameters

EA Parameter	Symbol	Value
Population Size	$\mathbf{P}^{(0)}$	300

8.5.10 Spartans

Testing of the Spartan concept on T_6 showed that its effects were detrimental to convergence to the point of failure. Previous testing on the klystron problem shows that this effect is not experienced. With this in mind the tuning of the number of Spartans was delegated to previous work on the single objective ellipsoidal problem.[16]

Table 8.14: Sweep of s simulation setup[16]

EA Parameter	Symbol	Value
Number of Spartans	s	$\{0, 20, \dots, 160\}$
Test problem	$\mathbf{F}(x)$	Ellipsoidal
Number of generations run for	G	10,000
Population Size	$\mathbf{P}^{(0)}$	100

Although this work is single objective and mono-modal there is still a convergence advantages to the use of Spartans. It will be assumed that the same advantages will be found for the klystron problem as testing can be considered to be excessively time consuming.

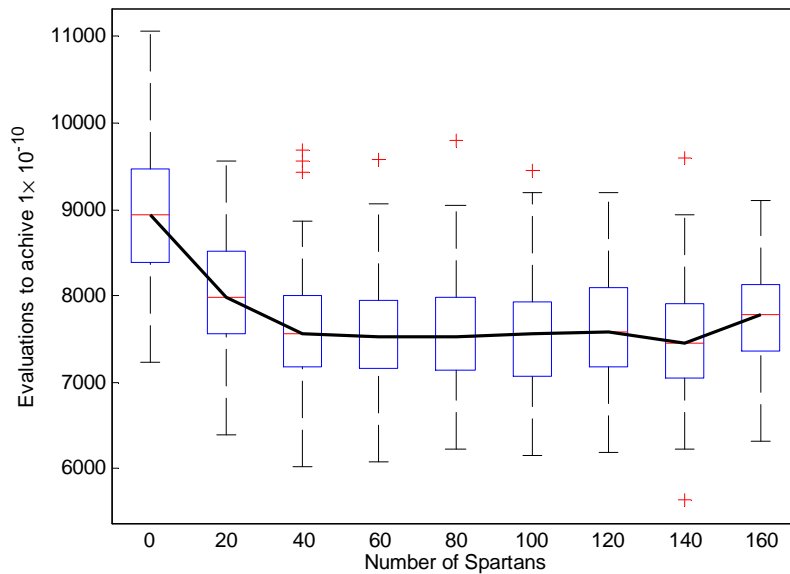


Figure 8.20: Single-objective number of spartan s sweep performed on T_6 . Showing that the optimal number of Spartans is around 40.

8.5.11 Population Management

The different methods of population limitation give little intuitive indication as to which will perform better. Three described above (see sec. 8.3.2) are evaluated against each other here:

Rej A fixed size population model which truncates the population after sorting by fitness (see sec. 8.3.2.1).

RejSig Another fixed size population model which truncates the population after sorting by “fitnificance” (see sec. 8.3.2.2).

Sig A maximum significance index population model (with a maximum significance index of 15) (see sec. 8.3.2.3).

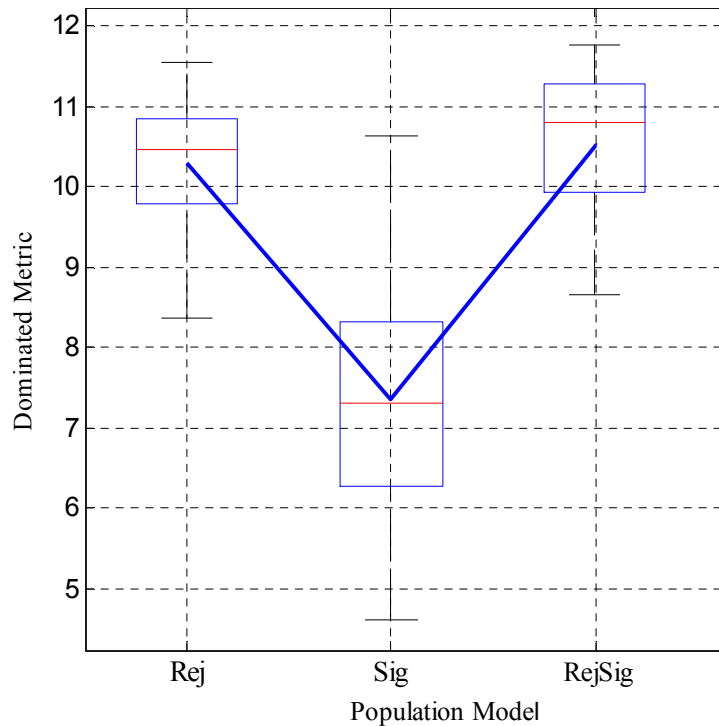


Figure 8.21: Comparison of population limitation models. This demonstrates there is a slight benefit in using rejection combined with “fitnificance”.

From fig. 8.21 there appears to be a slight advantage in using rejection in conjunction with “fitnificance”. (see fig. 8.21).

8.5.12 Calibration Conclusions

The most important parameters of the EA have been calibrated to clearly demonstrable values (see table. 8.15). The dominated metric clearly demonstrates the effect of each parameter change. Mutation has proved to be of little use, as its purpose is to add diversity, this goes some way to demonstrating that sufficient diversity has already been maintained by other means.

Although Spartans cause the test function used to fail, as initial test runs have shown they do not have the same effect when klystron simulations are used. An analysis of their utility for multi-objective optimisation is left for future work.

Table 8.15: Calibrated parameters

EA Parameter	Symbol	Value
Number of Spartans	s	0
Population management method		Reject (fitnificance)
Population Size	$P^{(0)}$	300
Mother Selection	$[\omega]_m$	Best in population
Parent Selection	$[\omega]_p$	Binary Tournament
Search strength	σ_x	0.8
Mutation Strength	σ_μ	0
Mutation Probability	$P(\mu)$	0
Number of generations	G	25,000
Number of Spartans	s	40
Spartan kill strength	σ_s	0.3

8.6 Klystron Optimisation Results

The klystron optimisation will first be performed using a Thales designed klystron TH2178 (see sec. 8.6.1); then the method will be applied to the improved TESLA specification (see sec. 8.6.2). Each klystron specification will be optimised using the same EA settings as described in the parameter calibration (see table. 8.15).

Table 8.16: Optimisation space definition

Vector	Components
Decision vector	[drift1, drift2, drift3, drift4, drift5, freq2, freq3, freq4, freq5, freq6]
Solution vector	[efficiency, length, Δ_{gain} , u_{slow}]

The Pareto Search Pressure Focusing (see sec. 8.2.1) settings are shown in table 8.17. The efficiency uses a variable cut off point of 0.7 of the maximum currently achieved. This allows the highest efficiencies to be targeted even if the current maximum efficiency is smaller than the final value. Absolute values are more appropriate for Δ_{gain} and u_{slow} as the desired values are easier to attain, the initial population often contains individuals which meet both criteria. The cut off for u_{slow} is set below the speed preferred by Thales so a range of values can be seen.

Table 8.17: Pareto Search Pressure Focusing settings

Variable	Setting	Absolute or variable
Efficiency	70%	of maximum at this iteration
Δ_{gain}	-2.01	Absolute
u_{slow}	$2 \cdot 10^7$	Absolute

Each optimisation is run over approximately 30k iterations, to ensure convergence although a much smaller number of calculations is likely to be sufficient.

8.6.1 TH2178 Comparison

In order to attempt to evaluate the performance of this method of optimisation it can be compared to professionally designed tubes. Thales have kindly provided a fully specified klystron interaction structure for this purpose (see table. 8.18).

8.6.1.1 TH2178 Specification



Figure 8.22: The TH2178 klystron[25]

TH2178 is a 800 kW, 500 MHz 6 cavity, single beam klystron with a second harmonic cavity. Its continuous wave specification has little effect on the interaction structure as it presents predominantly cooling issues.

Table 8.18: TH2178 klystron specification

Operating frequency (MHz)	500	Drive power (W)	40
Beam voltage (kV)	74	Beam diameter (mm)	30
Beam current (A)	17.2	Tunnel diameter (mm)	50
Peak output power (kW)	800	Gap Length (mm)	35
RF pulse length (ms)	CW	Number of cavities	6
Efficiency (%)	64	Number of 2 nd harmonic cavities	1
Perveance μK	0.85	Interaction Length (mm)	1580

When AJDisk attempts to simulate this structure it over estimates the efficiency by 10%. No cavity voltages are available for comparison, but this would indicate that the cavity interaction is being over estimated.

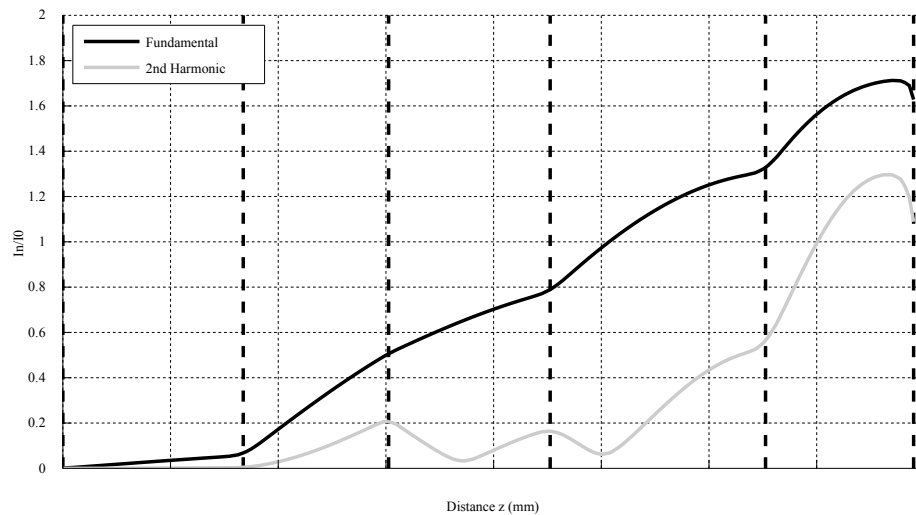


Figure 8.23: Modulation current plot growth for the fundamental and 2nd harmonic frequencies along the length of the klystron

Table 8.19: TH2178 AJDisk simulation results

Peak output power (kW)	942.472
Efficiency (%)	75.54
Reflected Disks	Yes

8.6.1.2 TH2178 AJDisk Calibration

AJDisk uses the longitudinal gap coupling factor averaged over the beam from 0 to r_b . Using the standard longitudinal gap coupling factor as calculated

by the analytical mode in AJDisk does not result in a simulation which matches the known tube performance. This is not completely unexpected as AJDisk ignores radial effects on the beam.

In order to correct this the simulation is tuned using the gap coupling factor until the output meets the specification. As the coupling factor for the fundamental frequency cavities is varied, the ratio between them and the 2nd harmonic cavity is kept constant.

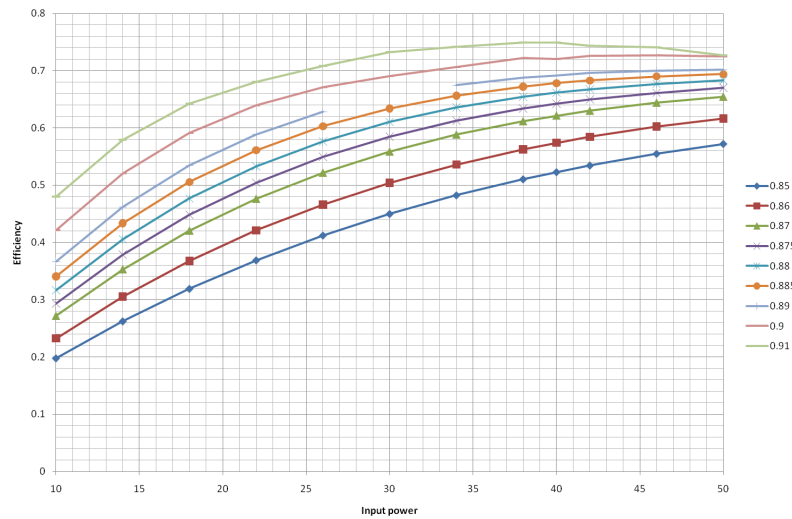


Figure 8.24: Calibration of AJDisk simulation using the gap coupling factor as the free parameter

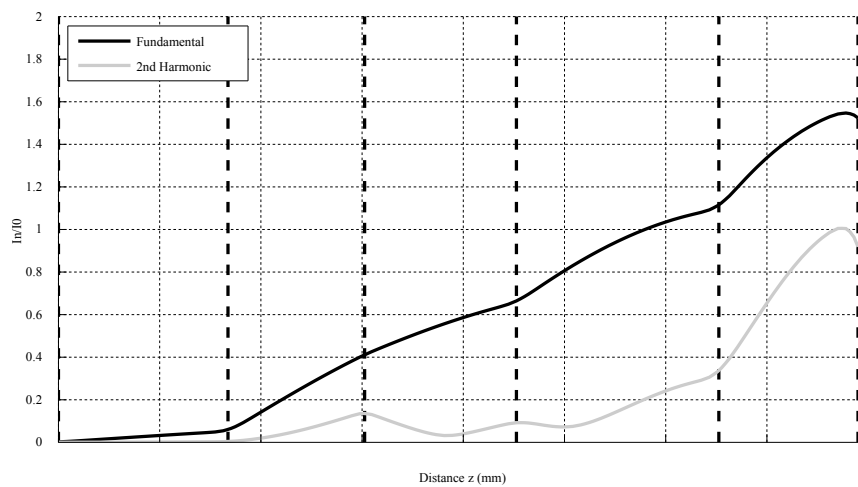


Figure 8.25: Modulation current plot growth for the fundamental and 2nd harmonic frequencies along the length of the klystron after calibration

The gap coupling factor which gives the most appropriate result (for 40 W input power) is 0.875 compared to AJDisk's calculated value of 0.9174. This does have electrons leaving the output gap slower than 8 kV ($5 * 10^7$) which could lead to reflection.

This work will assume that these simulations are accurate and these gap coupling factors are suitable for other interaction structures. However the optimisation has the caveat that, due to the wide variety of decision vectors which will be investigated and AJDisk's erroneous uncalibrated simulation, there is no guarantee the structures found are physical, or related to the TH2178. The major outcome of this optimisation will be proof of the optimisation process and trends due to alteration of variables. If this algorithm were applied to a more accurate code these issues would be eliminated as far as possible.

Table 8.20: TH2178 AJDisk post calibration simulation results

(a) Coupling factors

Cavity Gap	Calculated Gap coupling factor (z)	Calibrated Gap coupling factor (z)
Standard	0.9174	0.8750
2 nd Harmonic	0.9403	0.8969

(b) Simulation outputs

Peak output power (kW)	801.652
Efficiency (%)	64.25
Reflected Disks	yes

As all of the simulations for this klystron result in a u_{slow} below the reflection threshold, as for the purposes of this optimisation it has already been relaxed a range of u_{slow} will be found.

8.6.1.3 TH2178 Optimisation

The initialisation range of the decision vector is shown in table 8.21. Although these values bound the specification values, this is not compulsory. MrCX is capable, as many other recombination operators are, of exploring new

areas of decision space.

Table 8.21: TH2178 initialisation range

(a) Drift lengths

	Drift 1	Drift 1	Drift 1	Drift 1	Drift 1
Max	0.45	0.45	0.45	0.45	0.45
Min	0.2	0.2	0.2	0.2	0.2

(b) Frequencies

	Frequency 2	Frequency 3	Frequency 4	Frequency 5	Frequency 6
Max	550	1020	550	550	510
Min	480	970	480	480	490

Convergence The convergence of our main figure of merit, efficiency, is shown in figure 8.26. The maximum value of the efficiency shown on includes non Pareto individuals, meaning they may not meet the specification bandwidth or they may have reflected electrons. The maximum efficiency increases rapidly up to 5k iterations, the remaining iterations allow the algorithm to attempt to define good areas of the decision space.

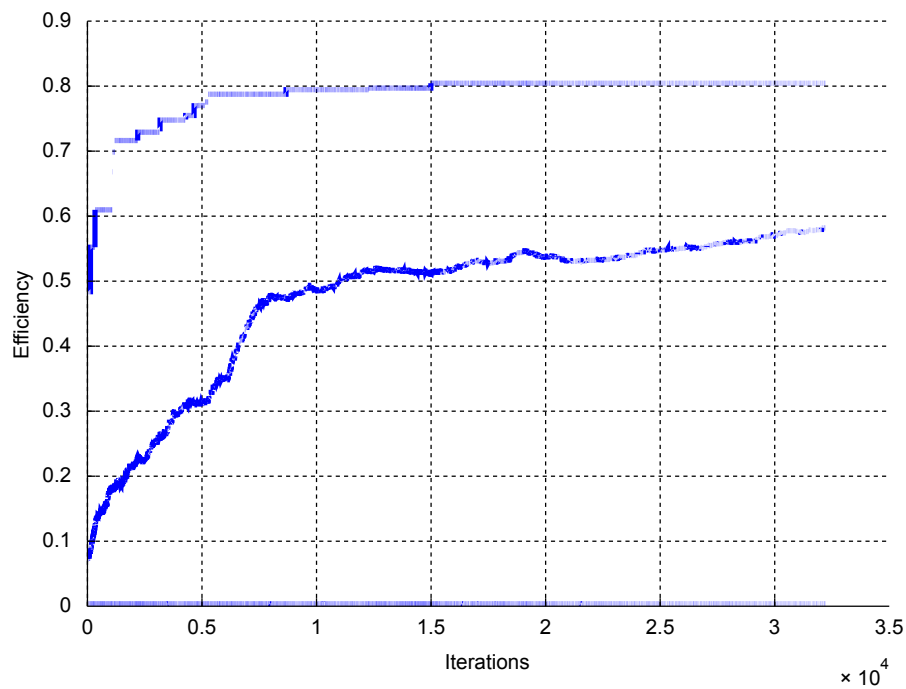


Figure 8.26: Convergence of efficiency against iterations. The top line shows maximum value and the middle line the average.

Drift Length The drift lengths have no common trends and can range over 250% of the specification value. The first two drift lengths have a clear optimum length which agrees with Thales specification. From the information available it looks as if a long drift after cavity 2 is not as detrimental as for drift 1 where efficiency rapidly drops off.

A short drift 3, which is immediately after the 2nd harmonic cavity, appears to lead to an inefficient klystron (with some exceptions). The specification drift length is slightly below a cluster of efficient klystrons and near individuals with higher u_{slow} . Increasing this drift appears to allow greater exit velocities while maintaining efficiency.

Drift 4 is showing evidence of multi-modality, more than one peak. The position of drift 5 is (as could be expected) critical, efficiency drops off markedly as the drift length increases. A strong relationship can be seen between efficiency (y axis) and u_{slow} (point size and colour map) although this is more clearly defined in fig. 8.29a.

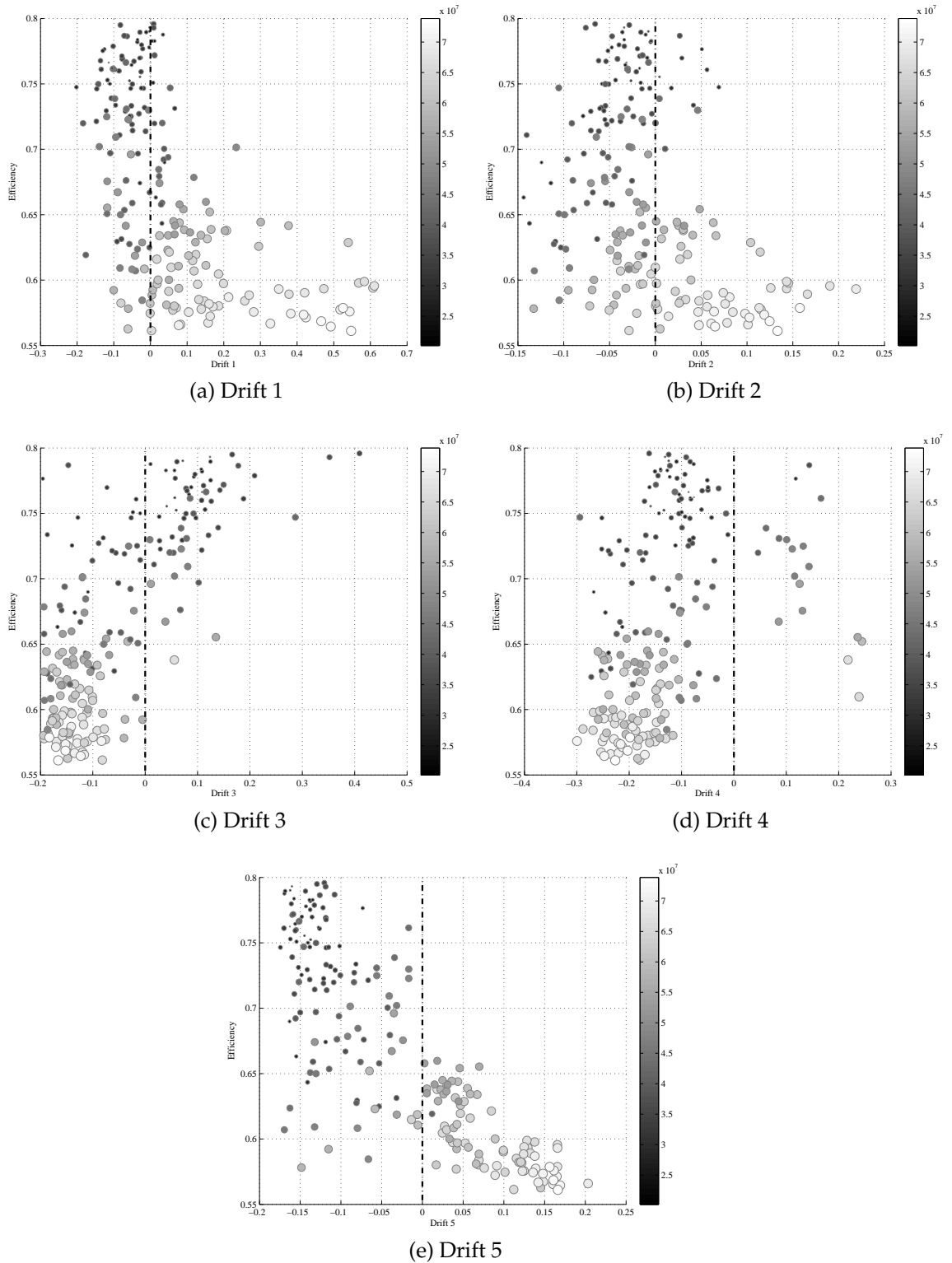


Figure 8.27: Deviation of the drift lengths (in meters) from the specification drifts in the final population against efficiency. Thales design drifts are marked with a vertical dotted line. The colour map and point size represent u_{slow} , bigger is faster.

Frequency The frequencies are defined within 5% of the optimum value, much more tightly defined than the drift lengths. Frequency 2 has been optimised to approximately the design value where frequency 4 and 5 differ from theirs by around 10 MHz (see 8.28a, 8.28c and 8.28d). Frequency 3, the 2nd harmonic cavity, converges above the specification frequency, this could be due to poor calibration of the gap coupling factor. Interestingly frequency 5 is the least defined (see fig. 8.28d), this could be due to the cavity frequency adjusting for the range of drifts observed (see fig. 8.27d).

Of special note is the output cavity frequency 6 (fig. 8.28e) is not set to the operating frequency (see fig. 8.28e). On average for frequencies above the operating frequency the electrons leave the klystron significantly faster than below. This could be due to approximations in AJDisk or it could indicate that, for this configuration at least, higher efficiency can be found using a slightly lower output frequency.

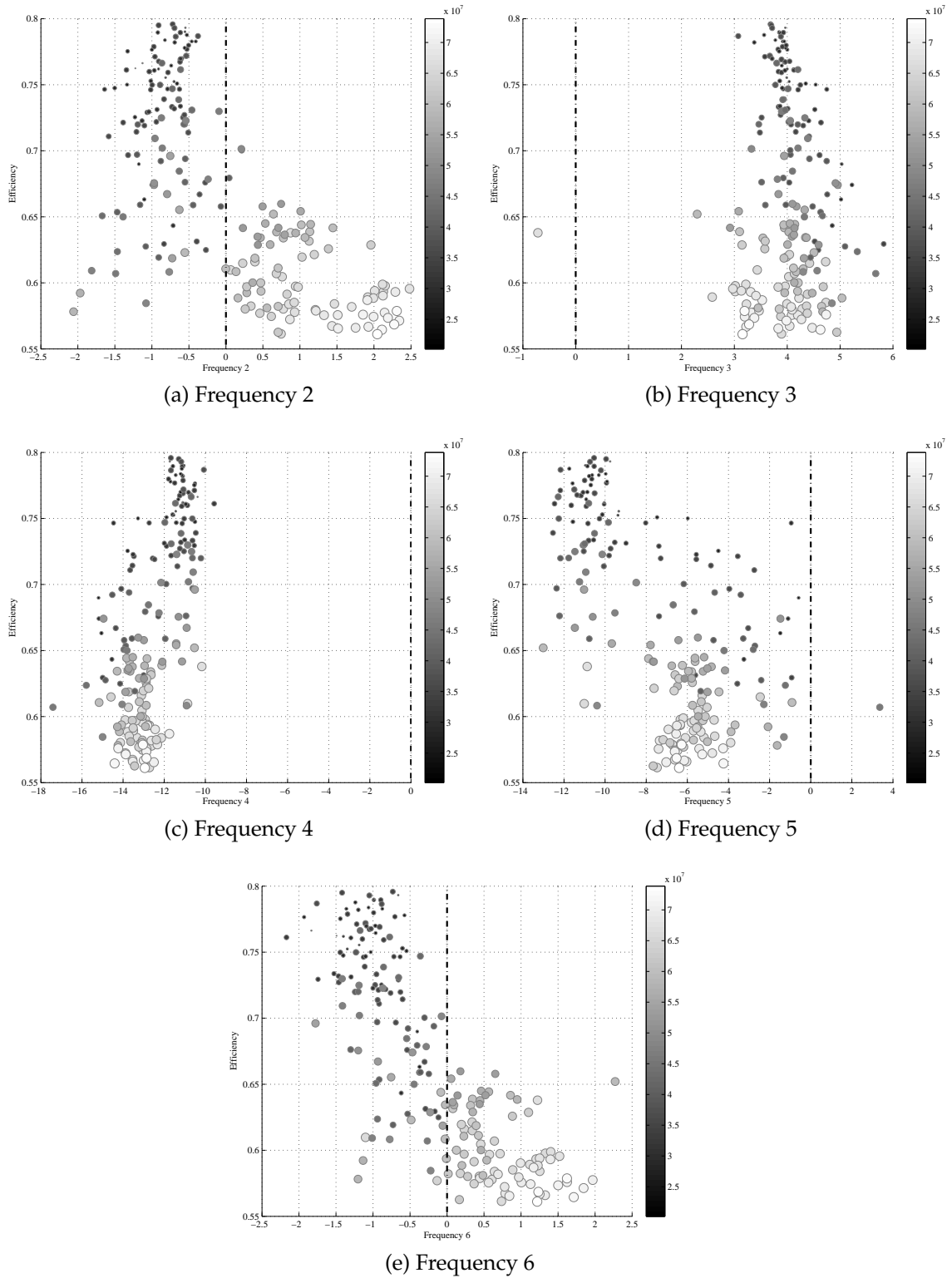


Figure 8.28: Deviation of the frequencies (in MHz) from the specification drifts in the final population against efficiency. Thales design frequencies are marked with a vertical dotted line. The colour map and point size represent u_{slow} , bigger is faster.

Solutions The relationship between exit velocity and efficiency can be clearly seen (see fig. 8.29a), it stands to reason that higher efficiency tubes (and by definition high output power) will slow the electrons more than an inefficient tube. The range of u_{slow} for a given efficiency shows that it is possible to spread the deceleration over the bunch, and a poorly optimised klystron will take too much energy from one part of the bunch. The larger, lighter points in fig. 8.29a represent shorter klystrons, there appears to be a near linear relationship between exit velocity and efficiency for klystrons of equal length.

Less clear is the definition of the relationship between length and efficiency. Around 60% efficiency the value of optimising the length is particularly clear. The shortest klystrons with the highest efficiency do define two clear limits. As the klystrons become shorter and more efficient, u_{slow} decreases (see fig. 8.29b).

Interestingly for a plot of u_{slow} against length (see fig. 8.29c), all the efficient klystrons clump at a certain length, 1.2-1.4m. A clear front appears in this plot showing that a short interaction length somehow limits u_{slow} for optimum klystrons.

PSPF applied to Δ_{gain} has worked well (see fig. 8.29d), there are no points lying below the target bandwidth (denoted by a Δ_{gain} of -2). As this solution is almost always -2 (unless the bandwidth is difficult to achieve) the optimisation collapses to 3 dimensions.

While the highest efficiency found (85%) is much higher than the 65% predicted by Thales for their specification, it must be remembered that the gap coupling factor may not be valid for all individuals.

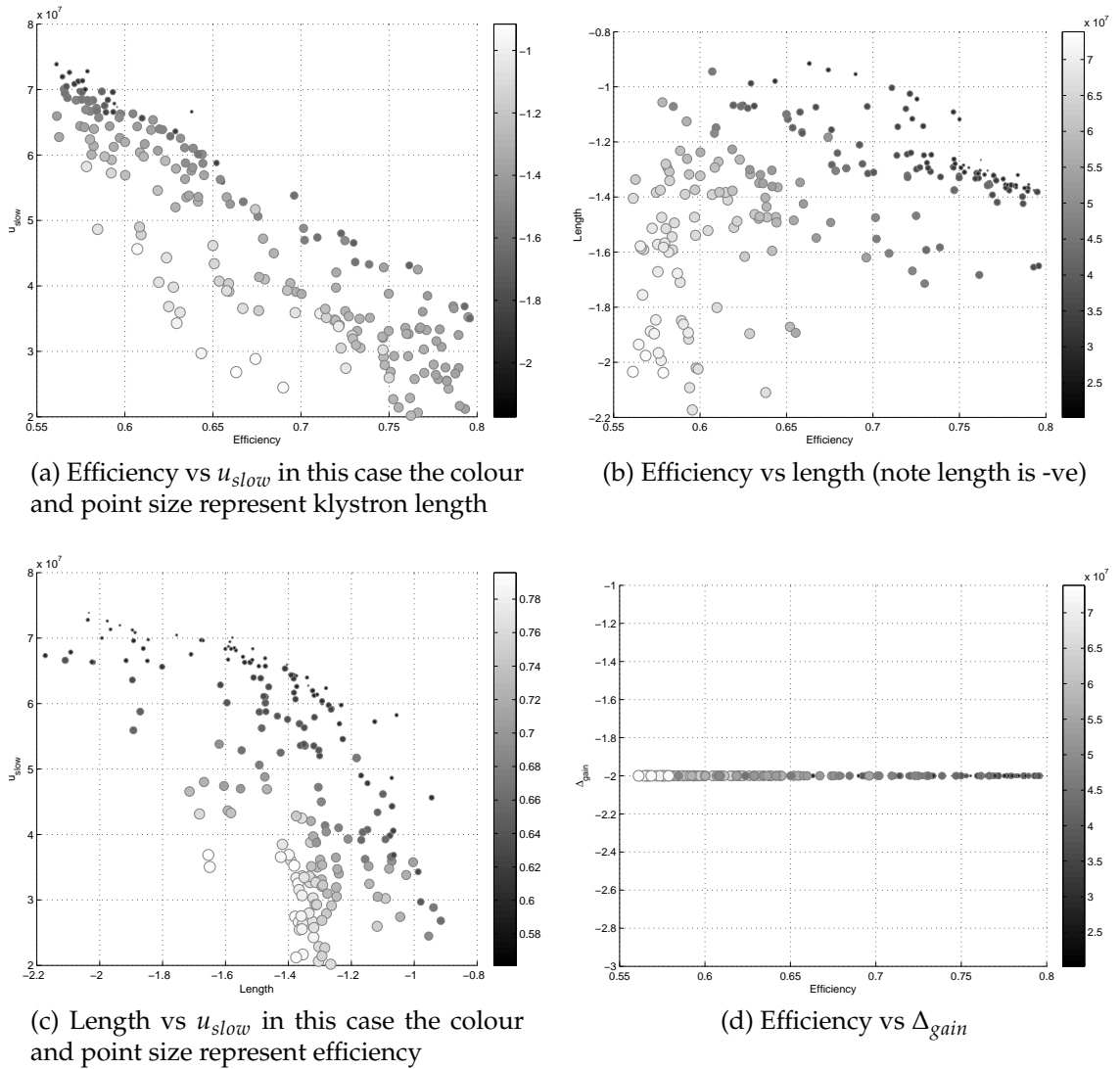


Figure 8.29: Plots of the relationships between the components of the solution vectors. Unless otherwise stated the colour map and point size represent u_{slow} .

8.6.2 Improved TESLA Optimisation

8.6.2.1 Improved TESLA Specification

The purpose of this algorithm is not to compare to existing designs but generate new ones. The TESLA specification combined with the cavity properties obtained from the cavity study was optimised by hand (see chap. 2). This interaction structure is used as the specification target for comparison.

The error in the current specification (8.21 A instead of 8.15 A) is repeated

here to ensure comparability (see sec. 6.2.1.2).

Table 8.22: Improved TESLA klystron specification

Operating frequency (MHz)	1300	Drive power (W)	200
Beam voltage (kV)	115	Beam diameter (mm)	3.5
Beam current (A)	8.21	Tunnel diameter (mm)	5.8
Peak output power (kW)	750	Gap Length (mm)	16
RF pulse length (ms)	1.5	Number of cavities	6
efficiency (%)	80	Number of 2 nd harmonic cavities	1
Perveance μK	2.11	Interaction Length (mm)	2.2

Table 8.23: Improved TESLA klystron hand optimised interaction structure (see chap. 2)

Cavity	drift (mm)	gap (2g) (mm)	F (MHz)	R/Q (Ω)	Q_0	Q_{ext}
1	0	16	1300.7	200	5000	300
2	0.6	16	1312.5	200	5000	
3	0.34	9	2544	150	3000	
4	0.508	16	1371	200	5000	
5	0.4	16	1339	200	5000	
6	0.22	16	1303	100	5000	60

Using AJDisk's calculated gap coupling factors the output power is under predicted (see table. 8.24). The predicted growth of the modulation current is shown in figure 8.30.

The modulation current is still rising at the output gap suggesting a longer drift length would have increased efficiency. This suggests that the cavity interaction strength is being under predicted leading to slow convergence of the bunches, or the space charge forces are being over predicted.

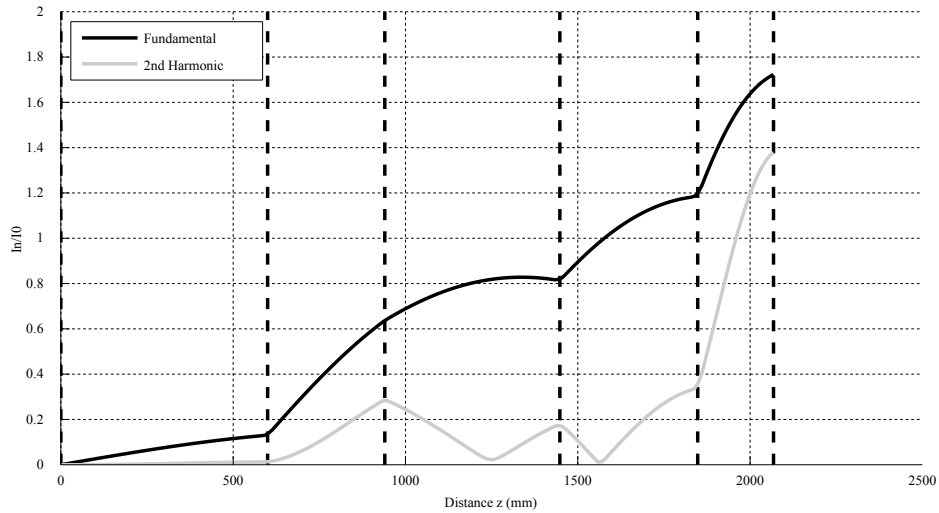


Figure 8.30: Modulation current plot growth along the length of the klystron

Table 8.24: Improved TESLA AJDisk simulation results

Peak output power (kW)	676.10165
Efficiency (%)	71.96
Reflected Disks	No

8.6.2.2 Improved TESLA AJDisk Calibration

In the absence of experimental results, AJDisk is calibrated to agree with Klys2D's results as Klys2D is a more complete and validated code. The gap coupling factor was optimised over a range of beam sizes around the specification of 3.5 mm. The optimal transfer curves for each beam size are plotted below. It can be seen that the specification beam size only yields a maximum efficiency of 75%. If the beam size is increased to 4 mm or 4.25 mm the target efficiency is achieved.

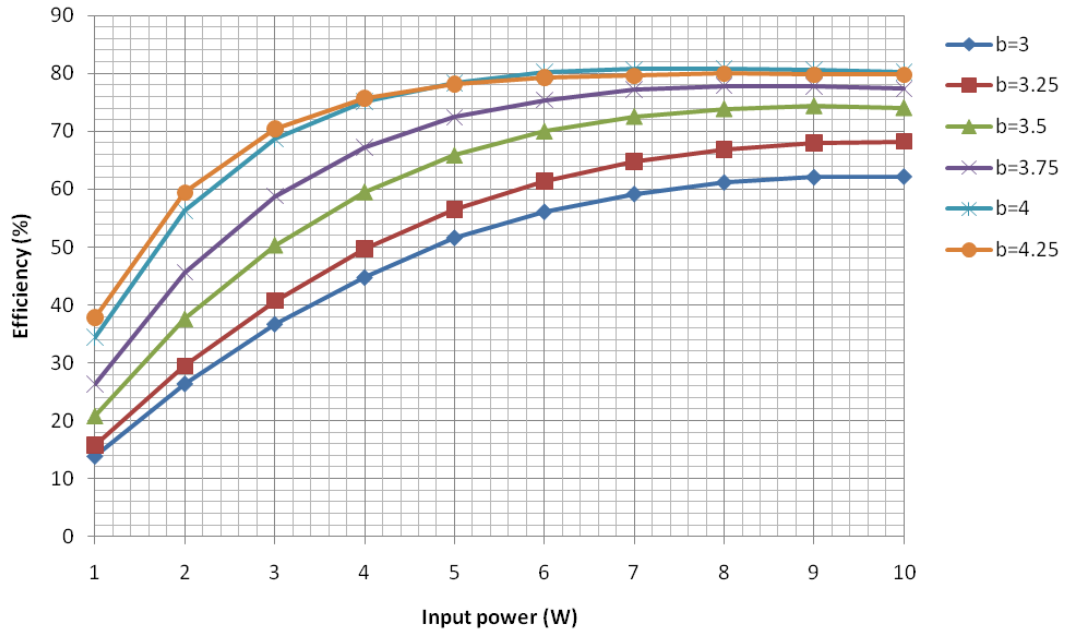


Figure 8.31: Calibration of AJDisk for a range of input powers (specification $P_{in} = 10W$) over a range of beam sizes with optimal gap coupling factors

Increasing the beam size lowers the space charge forces and the optimum gap coupling factor for these beam sizes is larger than the original calculated by AJDisk. This acts against the two predicted simulation issues above. The expected operating efficiency of 80% is achieved.

Table 8.25: Improved TESLA AJDisk post calibration simulation results

(a) Beam size

	Beam Size (mm)
Standard	3.5
Calibrated	4

(b) Coupling factors

Cavity Gap	Calculated Gap coupling factor (z)	Calibrated Gap coupling factor (z)
Standard	0.9539	0.9650
2 nd Harmonic	0.9777	0.9891

(c) Simulation outputs

Peak output power (kW)	754.17
Efficiency (%)	80.19
Reflected Disks	Yes

Compared to the Klys2D prediction of modulation current growth and voltage, a calibrated AJDisk over reports modulation current for cavities 3, 4 and 5. AJDisk also over reports the voltage by around 3% on cavities 1,4 and 5. The high voltage in cavity 3 could be explained by calculation of the voltage of a different harmonic. One may calculate the 2nd harmonic voltage and the other the fundamental. The agreement between the voltages in cavity 6 is no surprise as these are strongly linked to the output power which has been specifically tuned.

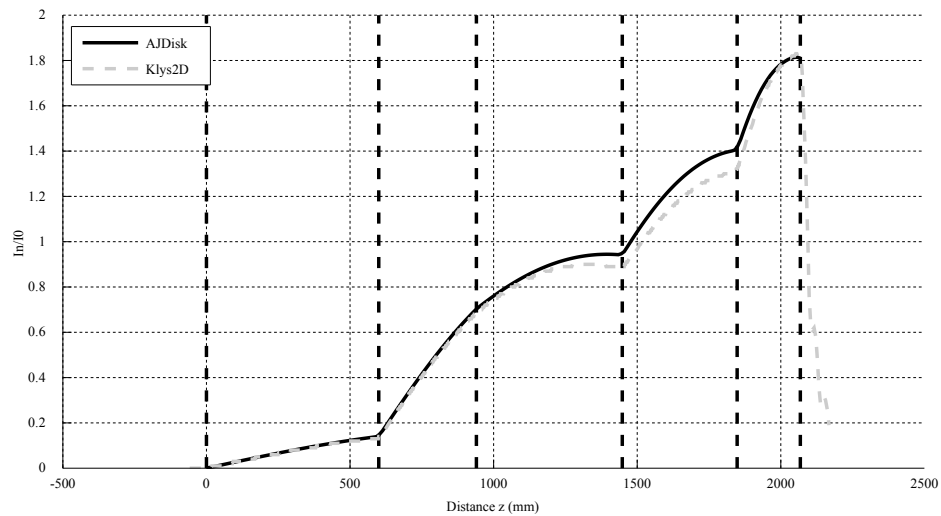


Figure 8.32: Comparison between AJDisk (black) and Klys2D (grey dashed)

Table 8.26: Comparison of gap voltages between AJDisk and Klys2D

	AJDisk (kV)	Klys2D (kV)	
1	1.8714	1.8	3.97%
2	11.9981	11.8	1.68%
3	6.7335	0.386	1644.43%
4	14.1055	13.6	3.72%
5	37.8779	36.7	3.21%
6	116.508	116.6	-0.08%

8.6.2.3 Improved TESLA AJDisk Optimisation

The initialisation range of the decision vector is shown in table 8.27 again this bounds the specification values.

Table 8.27: Initialisation range

(a) Drift lengths (m)

	Drift 1	Drift 1	Drift 1	Drift 1	Drift 1
Max	0.2	0.2	0.2	0.2	0.2
Min	0.7	0.7	0.7	0.7	0.7

(b) Frequencies (MHz)

	Frequency 2	Frequency 3	Frequency 4	Frequency 5	Frequency 6
Max	1200	2500	1200	1200	1200
Min	1400	2700	1400	1400	1400

Convergence The convergence of our main figure of merit, efficiency, is shown in figure 8.33. The highest efficiency is found after only ≈ 1000 calculations, and the remaining evaluations are again used to explore the decision space. This optimisation, at least from an efficiency point of view, has converged much more completely than the TH2178. The maximum efficiency is quite stable, and the mean is not rising rapidly. The fluctuations in the minimum are the result of simulation vetoes, most probably unconverged simulations, which are marked with an efficiency of 0% (see sec. 8.2.2).

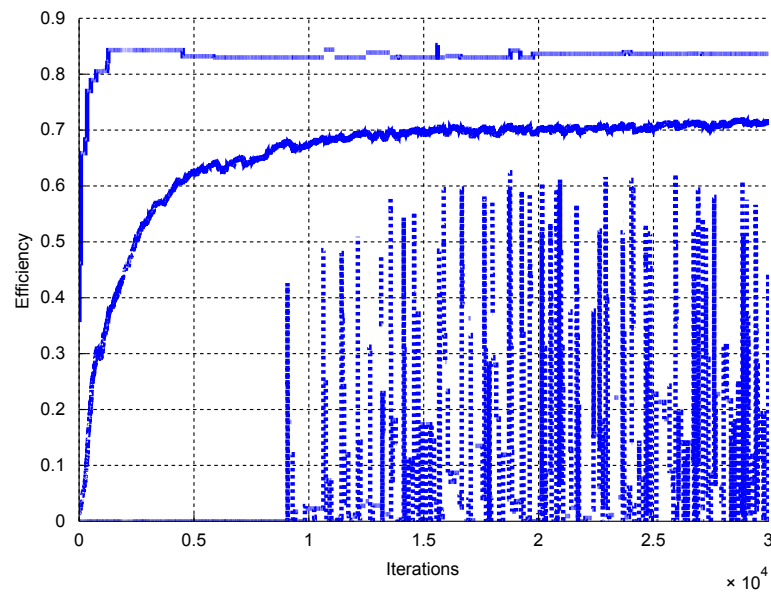


Figure 8.33: Convergence of efficiency against iterations. The top line shows maximum value and the middle line the average and the bottom the minimum.

Drift Length The drift lengths in this optimisation show less correlation to efficiency than for this TH2178, although klystrons with higher u_{slow} do tend to clump around certain values. A range of almost half a metre covers the ranges for most of the drifts. The exception is drift 5, a strong drop in efficiency and increase in u_{slow} can be seen for increasing drift length (see fig. 8.34e). Drift 5 also has the best correlation between optimisation and the hand optimised value. Although the remaining hand optimised values are within the large ranges. Again it is clear that efficiency and u_{slow} are strongly related.

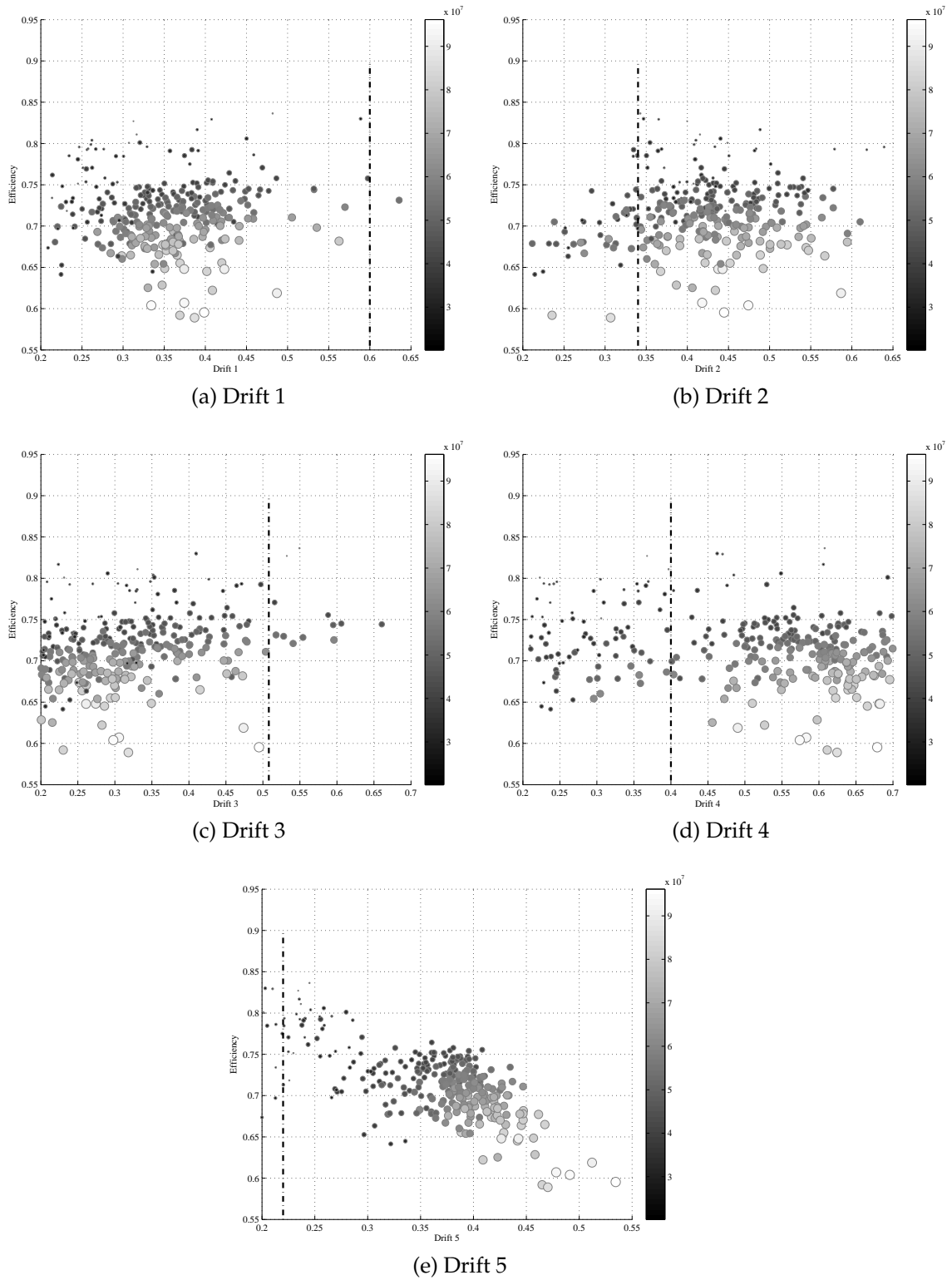


Figure 8.34: The drift lengths (in meters) of the individuals in the final population vs efficiency. Hand optimised drifts are marked with a vertical dotted line. The colour map and point size represent u_{slow} , bigger is faster.

Frequency The highest density of points appears to be at lower efficiencies and is clumped for each frequency. This density could be an artifact of the convergence of the algorithm or it could be a real property of the search space. It should be remembered that each point represents the result of 10 decisions not only one. If small changes in any of these values lead to a large change in one of the solutions this sparse plot could result. Clustering is only applied in the solution space and does not have a similar effect on decision space.

The smallest range is again the output frequency, for this specification the highest efficiency is found slightly above the operating frequency. u_{slow} again decreases rapidly for lowering frequency.

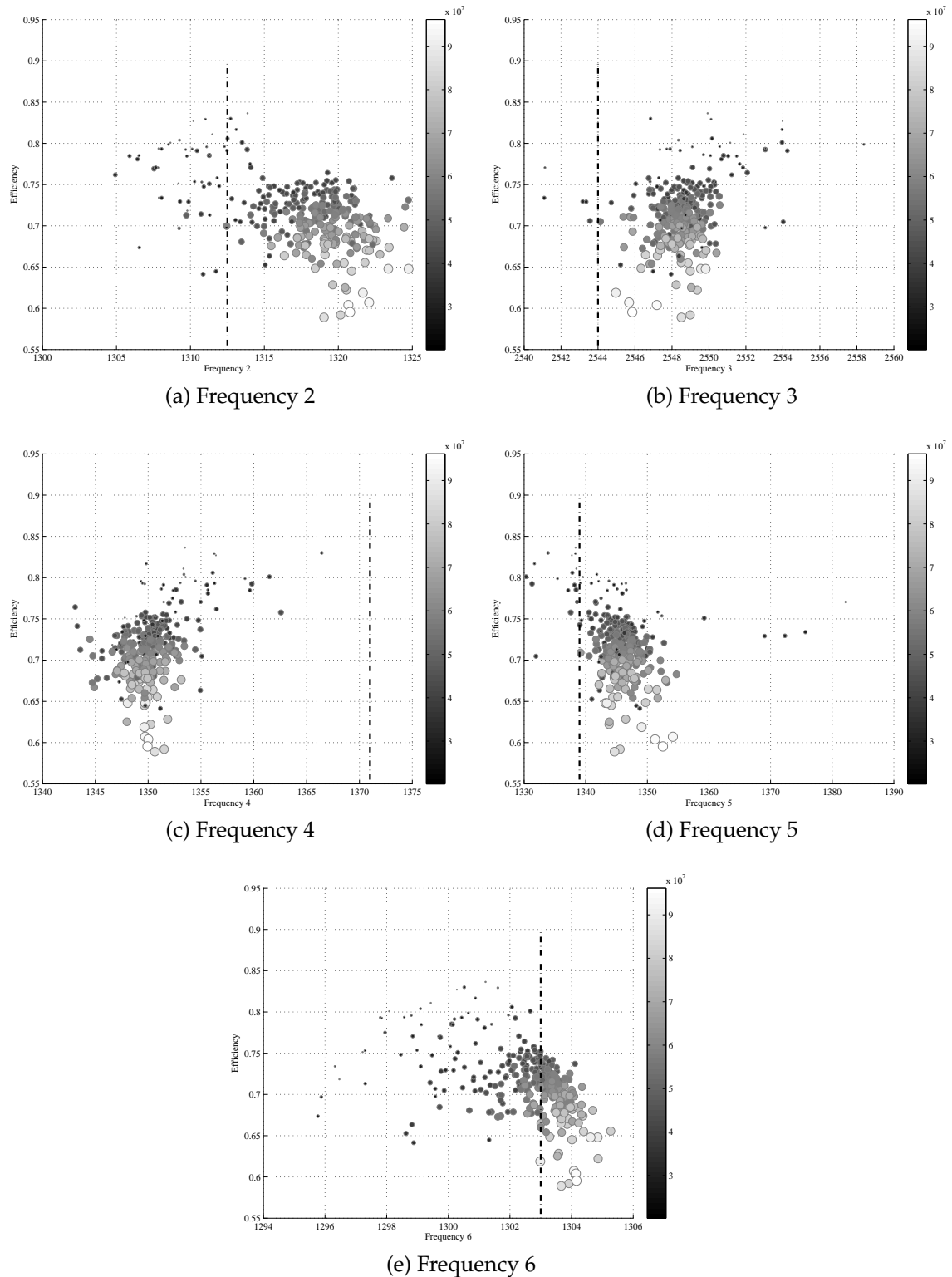


Figure 8.35: The frequencies (in MHz) of the individuals in the final population. Hand optimised frequencies are marked with a vertical dotted line. The colour map and point size represent u_{slow} , bigger is faster.

Although the optimised values for drift and frequency do not agree with the hand optimised values as well as for the TH2178, this is likely to be an

indication of the author's skill at optimising interaction structures. It must be remembered that if a klystron is found with equal efficiency but a shorter length (and so different drifts and frequencies) the longer klystron will be considered dominated and its decision space avoided.

The limit on efficiency imposed by klystron length is clearer from this optimisation as is the limit imposed by exit velocity of electrons. The knee is just visible at the top of the plot, around $9 \cdot 10^7$ m/s. The limit imposed by u_{slow} appears to be linear for at least a significant range and again holds for klystrons of equal length. However it appears that for lower u_{slow} the linear relationship is broken, and it is known that this is the case for higher u_{slow} .

Although the front is indistinct there is a relationship between length and efficiency, with efficiency dropping rapidly for decreasing length. This relationship appears to hold for structures of constant u_{slow} . There appears to be an outlier on the length-efficiency curve. This could be due to its low u_{slow} ; indicate that there are areas of decision space that the algorithm has not investigated or it could indicate an anomalous result from AJDisk.

Again PSPF has ensured that klystrons that meet the specification bandwidth are preferentially selected. There are individuals which do not have a Δ_{gain} of exactly -2 due to the PSPF cutoff of -2.01 but they are within the specification to 2 d.p.

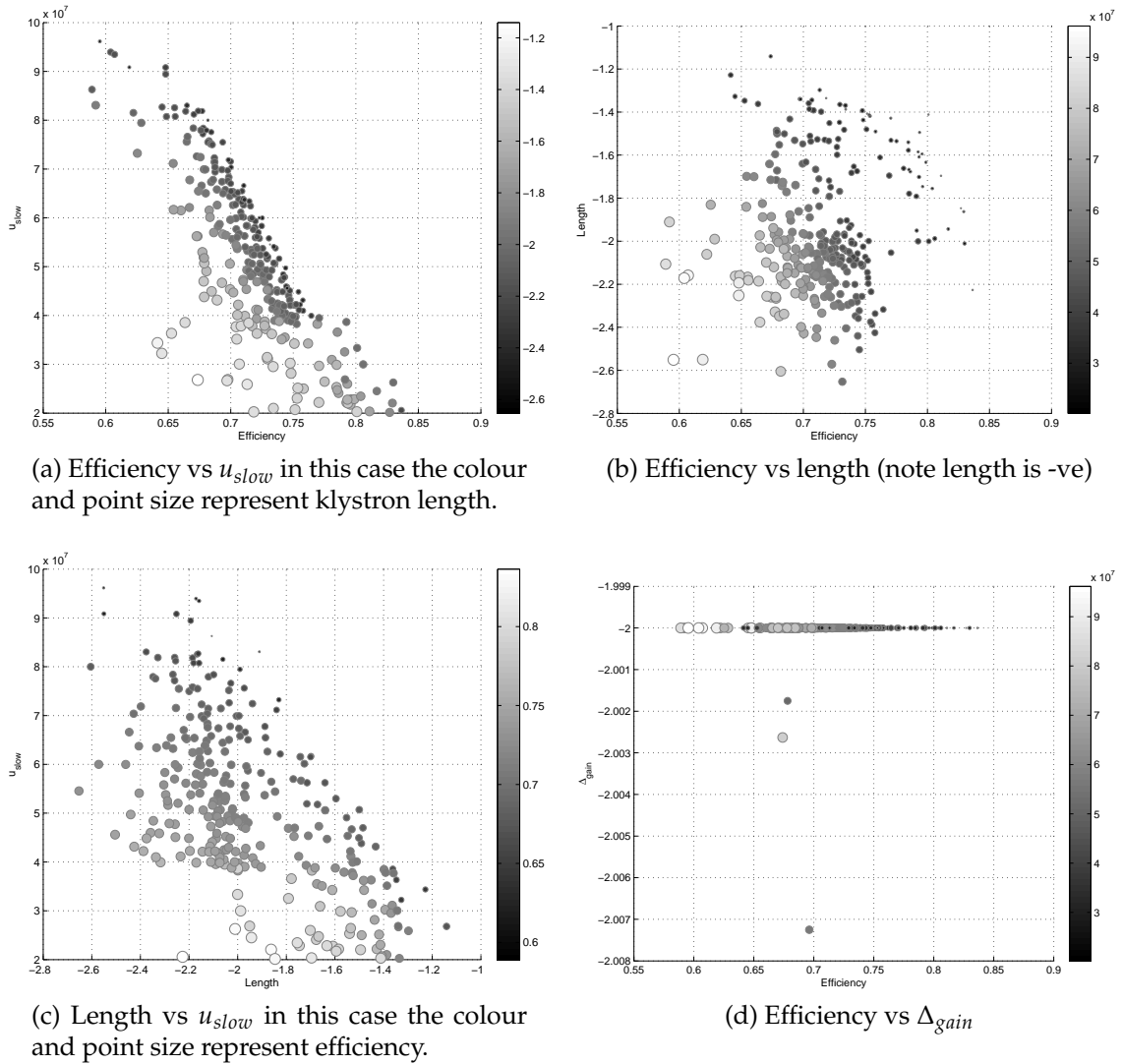


Figure 8.36: Plots of the relationships between the components of the solution vectors. Unless otherwise stated the colour map and point size represent u_{slow} .

8.7 Conclusions

The manual optimisation of a klystron interaction structure is a slow, time consuming task. This is a task that can be automated by the use of an evolutionary algorithm. As there are a number of objectives (length, efficiency, Δ_{gain} and u_{slow}) a multi-objective algorithm must be used. This allows all of the target properties of the klystron to be optimised simultaneously.

Using AJDisk klystron simulation software combined with a Condor pool,

thousands of interaction structure evaluations can be performed quickly and automatically. The poor performance of AJDisk before calibration on the test problems has limited the conclusions drawn to validation of the optimisation methods and the observation of trends. Any quantitative conclusions, such as maximum efficiency obtainable, cannot be relied upon. However, with the simulation tool used, the algorithm successfully optimised the interaction structures to a high efficiency while maintaining a short length, with a high electron exit velocity.

A number of techniques are required to aid compatibility between the optimisation approach and the algorithm used. For instance Pareto Search Pressure Focusing allows only efficient klystrons to be targeted, as opposed to klystrons short but inefficient. This approach has been extended to the issue of reflected electrons. The speed of the slowest, and so most likely to be reflected, electron is maximised along with the other objectives, with preference given to those which will not reflect using PSPF. A new definition of bandwidth (Δ_{gain}) must also be used to avoid excess simulations.

Interesting properties of the solutions space become visible. For instance the speed of the slowest electron when plotted against efficiency has a pronounced knee, showing 2 distinct regimes of output gap operation.

Pareto fronts have formed in the solution vector plots. These are not easily observed as they are 3 dimensional in nature. They do however reveal a near linear relationship between efficiency and exit velocity in the region of interest and curves the efficiency-length and efficiency- u_{slow} plots.

Of particular interest in a commercial setting is the speed of convergence of the algorithm. Excessive evaluations (as used here) are not possible when access to computing resources is limited. It is likely that a skilled engineer could outperform this algorithm, albeit with a single solution rather than many. Further research could be undertaken into techniques which allow a klystron design to be reached within fewer evaluations.

Chapter 9

External Coupling

9.1 Introduction

The couplers are the link between the klystron and the low level RF source and load. Their correct operation ensures that the RF energy is effectively transmitted to the beam at the input gap, then removed from the tube at the output.[34]

A well designed coupler couples strongly to the desired mode, has of course an appropriate Q_e for the operating mode and allows the cavity to maintain an appropriate resonant frequency.[2] In the case of an MBK it must also retain the field uniformity between the beams to avoid asymmetric modulation/demodulation.

Any drop in the focusing field in an MBK will cause a defocusing radial component to enter the field. The output waveguide often exits the tube perpendicularly to the beam direction through the solenoid. This can be compensated for by adjustments to the solenoid or by introducing additional coils, but if the output coupler can be designed to avoid this then remediation can be avoided.

The couplers proposed for the output cavity are derived from those used by Calabasas Creek Research, Inc. in a 30 MW gyro-klystron [26] (see fig. 9.1). Although the proposed coupler designs are structurally similar to those in the gyro-klystron where the coaxial line is coupled to E_r of a TE mode in the cavity, in the design proposed the coaxial line couples to the H_θ in a TM mode.

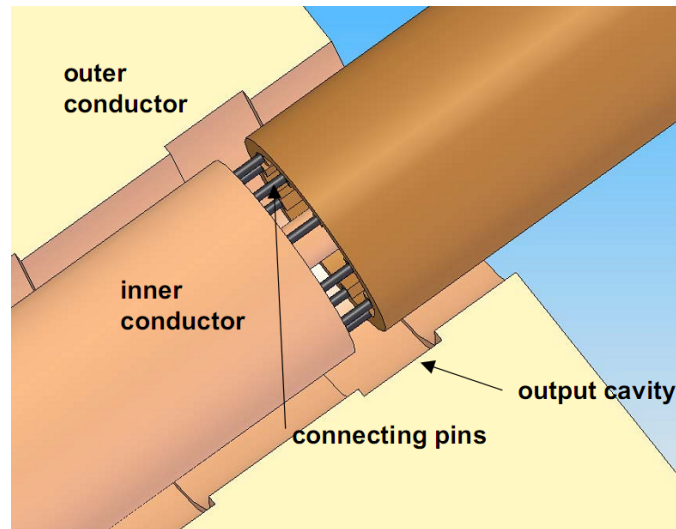


Figure 9.1: Gyro-klystron output coupler design used by Calabasas Creek Research, Inc.[26]

In the design shown in fig. 9.1 the connecting pins are conductive and perform a similar function to the inside wall of the coaxial cavity while still allowing the fields to couple through to the coaxial line. This has the advantage of helping to maintain rotational symmetry, although their periodic nature causes the field to be slightly uneven at the beam radius.

The design proposed here is largely unchanged in structure for the output coupler while the input coupler has small changes. The coaxial line is terminated to create an additional resonant coupled resonant cavity into which a much smaller lower power coaxial line is coupled.

The designs proposed have not been tuned to their final specifications, but their simulation has suggested that this should be possible. They do however highlight a number of problems in the design which must be addressed.

9.2 Coaxial Output Coupler

The output coupler is responsible for coupling power out of the output cavity into some form of high power waveguide. Often this may be just be a window on the side of the output cavity which links in to a piece of, often half

height, rectangular waveguide.[34] In the case of the proposed design to avoid perturbing the magnetic field the waveguide will exit the tube axially.

Table 9.1: Output coupler specification

Frequency	1.303 GHz
R/Q	$\approx 150 \Omega$
Q_e	≈ 60
Power	15 MW

Circular waveguide cannot be used to exit the output cavity as it is cut off for 1.3 GHz for TM_{01} and TE_{01} at the sizes compatible with the cavity. The cut-off frequencies can be calculated as follows [72]:

$$f_{cTM01} = \frac{0.383}{r\sqrt{\mu_0\epsilon_0}} = 1.531 \text{ GHz} \quad (9.1)$$

$$f_{cTE01} = \frac{0.609}{r\sqrt{\mu_0\epsilon_0}} = 2.434 \text{ GHz} \quad (9.2)$$

where $r = 75 \text{ mm}$, the largest feasible circular waveguide for this configuration.

TE_{11} is not cut off, but is also not rotationally symmetric [72]:

$$f_{cTE11} = \frac{1.841}{2\pi r\sqrt{\mu_0\epsilon_0}} = 1.171 \text{ GHz} \quad (9.3)$$

Coaxial waveguide allows for much lower cut off frequencies and for a piece of 50Ω coaxial line with the same outer dimensions.

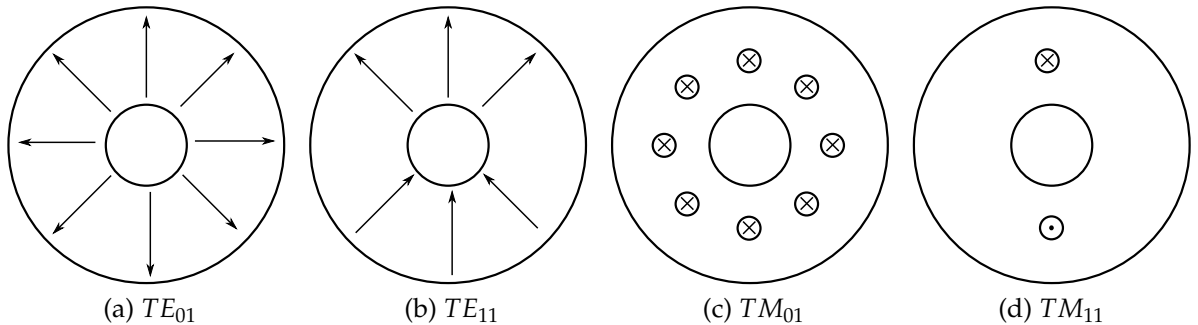


Figure 9.2: Electric fields of the modes supported by coaxial transmission line[27]

From the modes supported by coaxial transmission line (see fig. 9.2) the

asymmetric ones can be immediately dismissed (TE_{11} & TM_{11}).

Table 9.2: Coaxial mode cut-off frequencies

TE_{01}	TM_{01}
$f_{c TE01} = \frac{ck_c}{2\pi\sqrt{\epsilon_r}} = 0.887 \text{ GHz}$ [27] where $k_c \approx \frac{2}{d+D}$	$f_{c TM01} = \frac{c}{2.093(D-d)} = 1.68 \text{ GHz}$ [89]
where $D = 2r$ and $d = D/2.303$ (as explained in eq. 9.4)	

The cut off frequency for the TEM mode, $f_{c TEM}$ in coaxial line is 0 GHz. The power handling capability is the only constraint on coaxial line size. In an ideal world all other modes would be cut off so the size of the coaxial line could be adjusted to achieve this within the other constraints.

The coaxial line is matched to 50Ω using the equation [72]:

$$Z_0 = \frac{1}{2\pi} \sqrt{\frac{\mu_0}{\epsilon_0}} \ln \frac{D}{d} \approx \frac{138\Omega}{\sqrt{\epsilon_r}} \log_{10} \frac{D}{d} \quad (9.4)$$

where D is the outer diameter and d is the inner diameter yields the result that the ratio $\frac{D}{d} = 2.303$ for a 50Ω characteristic impedance.

The power handling capability of coaxial line of this size is defined by:

$$P_{coax} = \frac{E_0^2 \cdot d^2}{\zeta_0} \pi \ln \left(\frac{D}{d} \right) \quad (9.5)$$

where $\zeta_0 = 377 \Omega$ the impedance of free space and E_0 is 3 kV/mm[27]

Which in the case of a coaxial line with an outside diameter of 150 mm and a characteristic impedance of 50Ω gives a maximum power of 66.353 MW, which is well above the peak power of the tube 15 MW.

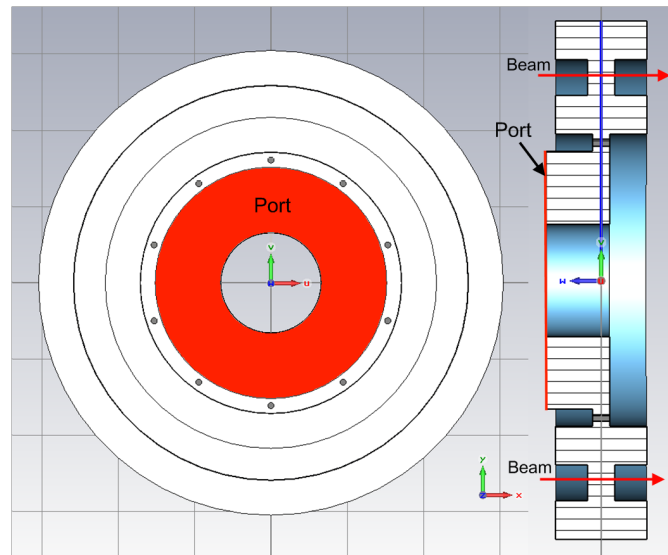
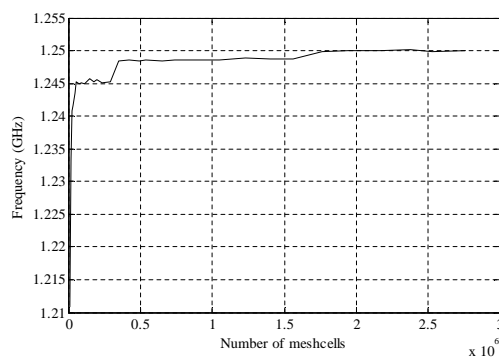


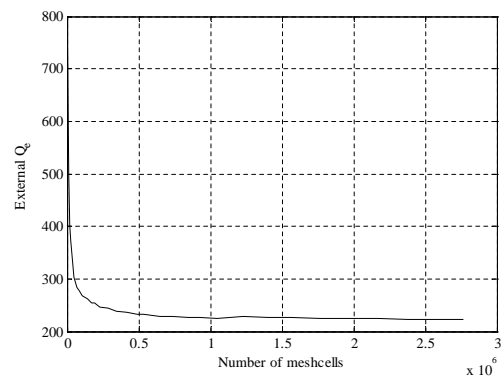
Figure 9.3: “Solid” vacuum representation of the output coupler with the ports marked

The coupler consists of the specified piece of coaxial line slotted inside a standard coaxial cavity. In this case using the same dimensions as defined in the cavity study for simplicity (see chap. 2), a 10 mm gap connects the two vacuum regions containing 20 2 mm radius rods modelled as perfect electric conductor (see fig. 10.2a).

9.2.1 Convergence



(a) Frequency convergence of the structure with respect to number of mesh cells

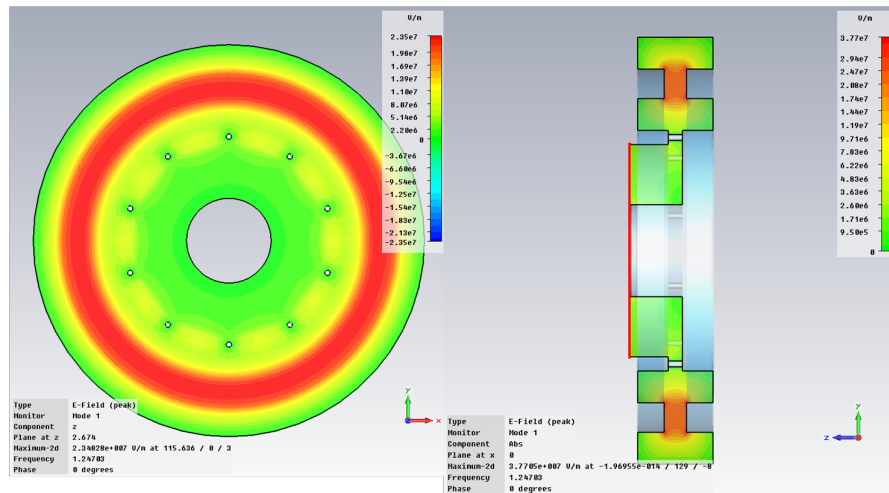


(b) Q_e convergence of the structure with respect to number of mesh cells

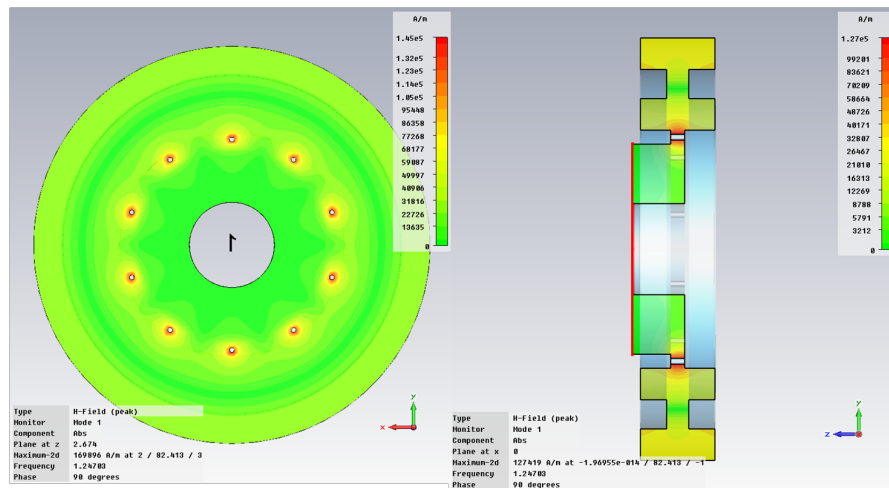
Figure 9.4: Convergence behaviour of the cavity

The number of mesh cells which represents a convergence ($2 \cdot 10^6$) corresponds to a value of 90 for the “mesh cells per wavelength” parameter in MWS. Using the determined mesh density the structure was simulated for the operating mode and the dipole mode.

9.2.2 Simulations



(a) $|E|$ field



(b) $|H|$ field, note the high field around the conductive posts.

Figure 9.5: Diagrams of the fields in the output cavity and coupler for the operating mode

Table 9.3: Operating and dipole mode properties (Q_e are for cavities with 10 posts)

Mode	Operating	Dipole
Frequency	1.245 GHz	1.316 GHz
R/Q	177.2 Ω	335.0 Ω
Q_o	7734.8	$1.126 \cdot 10^{10}$
Q_e	224.24	8066.6

In terms of fitting the specification of the input cavity this particular geometry needs optimisation. The frequency and Q_e both differ appreciably from their targets, although they could conceivably be tuned by altering the cavity dimensions.

In order to achieve a suitable order of magnitude for Q_e , the number of posts has been reduced to 10 from the 20 which would match most easily with the beams. This changes the relationship between beams and posts from a symmetry point of view and is the source of the majority of the ripple observed in E_z .

Observing the E_z magnitude on a plot at the beam radius in the cavity shows a good uniformity of field for each beam with a ripple of 0.193% (see fig. 9.6). The ripple would lead to differences between the beams as they interact with the cavity, some of them being non-optimal. Intelligent positioning of the beams in relation to the posts could reduce the ripple to an even lower level by positioning the beams at the average E_z points.

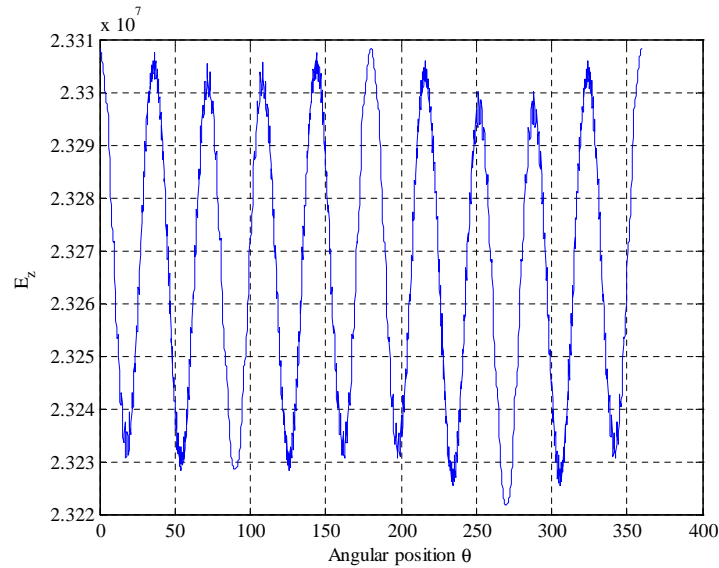


Figure 9.6: Plot of E_z around a circle passing through the beam centres. Showing the ripple caused by the conductive posts.

Critically, in this design, the plot of $|H|$ shows a high field concentration at the conductive posts which cannot be enlarged without reducing their number to maintain the correct Q_e . This will lead to a high degree of localised heating exacerbated by the small radius posts which will not accommodate sufficient cooling channels (see fig. 9.5b).

The frequency separation between the operating mode and the dipole mode has been little affected by the addition of the coupler with the separation at approximately ≈ 50 MHz (see fig. 9.7). The dipole mode does not couple out with a Q_e of 10^{10} reducing the impact on the output signal (see table. 9.3). However as this mode is trapped in the cavity there is the risk that it will build up to problematic levels over a short time and disrupt the beams.

The coupling posts prevent use of the dipole shifting method employed for the intermediate cavities, as if they are cut they no longer perform their function (see sec. 3.4), a different solution to this issue is required.

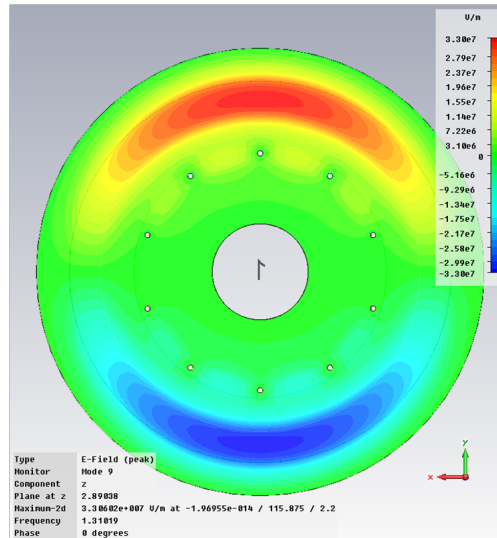


Figure 9.7: E_z of the dipole mode in the output cavity. It's frequency remains within ≈ 50 MHz of the operating mode

9.3 Input coupler

Table 9.4: Input coupler specification

Frequency	1.3 GHz
R/Q	$\approx 200 \Omega$
Q_e	≈ 300
Power	200 W

The input coupler is responsible for coupling low level RF into the input cavity where the beams can be modulated. In the case of the input coupler there is no advantage in the RF input entering axially, in fact this would obstruct the high voltage components in the region of the guns. In order to retain the advantages of the output coupler, the coaxial waveguide can instead be terminated to create a further coaxial resonant cavity. Coupling to the coaxial line however can add an asymmetry which will be observed later.

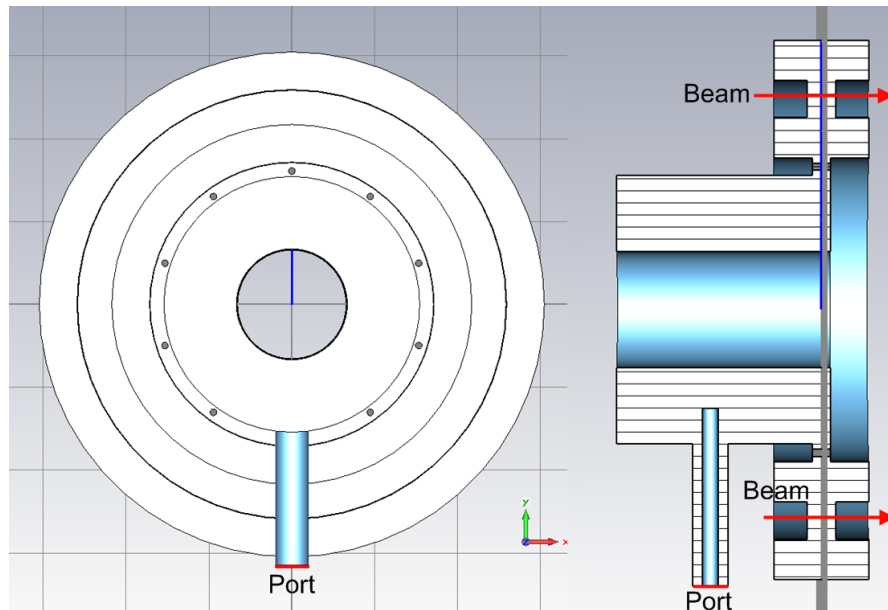


Figure 9.8: “Solid” vacuum representation of the input coupler showing the E probe coupling to a short resonant length of coaxial waveguide

In order to simplify the simulations, the cavity dimensions are the same as those chosen by the cavity study (see sec. 2.8) and for continuity a coaxial waveguide of identical dimensions is used. The length L of the resonant coaxial line is set to half a wavelength [72]:

$$L = \frac{v\pi}{\omega} \quad (9.6)$$

where $\omega = 2\pi f$ and $v = c$ in a vacuum [72]:

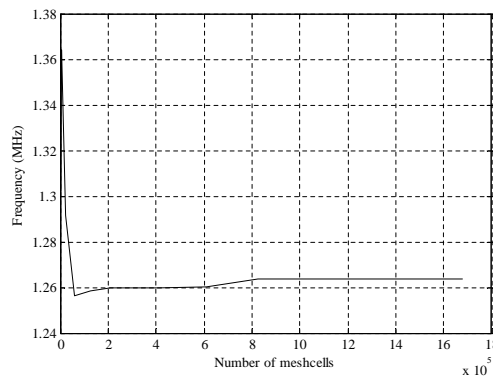
$$v = \frac{c}{\sqrt{\epsilon_r \mu_r}} \quad (9.7)$$

This resonant coaxial line of 1.3 GHz has a half wavelength of 115.3 mm.

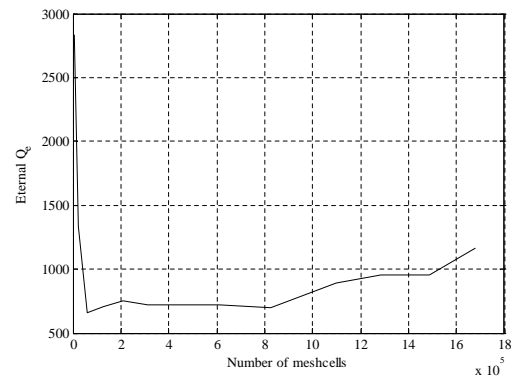
The input coaxial line has an arbitrary outside diameter of 10 mm and is also tuned to 50Ω . The length of the e-probe and the size of the gap containing the 2 mm radius conductive rods (again modelled as PEC) was coarsely tuned to achieve a suitable Q_e for the operating mode.

9.3.1 Convergence

The convergence of the MWS simulations was investigated with respect to frequency and Q_e . The former, being a first order property, would be expected to converge more quickly and the latter, a second order property, would be expected to require higher mesh density.



(a) Frequency convergence of the structure with respect to number of mesh cells

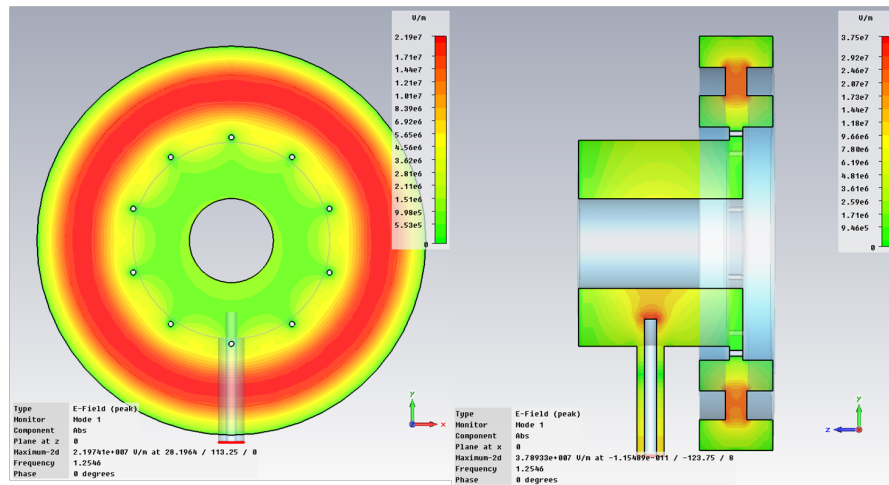


(b) Q_e convergence of the structure with respect to number of mesh cells

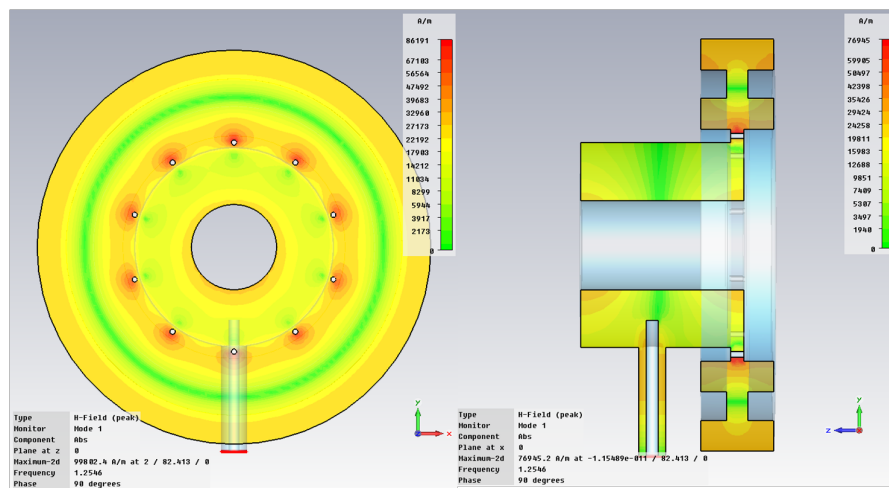
Figure 9.9: Convergence behaviour of the cavity

The number of mesh cells which represents a convergence ($1.24 \cdot 10^6$) corresponds to a value of 50 for the “mesh cells per wavelength” parameter in MWS. Using the determined mesh density the structure was simulated for the operating mode and the dipole mode.

9.3.2 Simulations



(a) $|E|$ field



(b) $|H|$ field, note the high field around the posts

Figure 9.10: Diagrams of the fields in the input cavity and coupler for the operating mode

Table 9.5: Operating and dipole mode properties

Mode	Operating	Dipole
Frequency	1.262 GHz	1.313 GHz
R/Q	157.8 Ω	325.9 Ω
Q_o	9328.1	8040.8
Q_e	701.66	7006.19

Again the cavity has not been fully tuned but adjustment of the frequency and Q_e are both feasible.

In order to achieve a similar order of magnitude for Q_e the number of posts

has been reduced to 10 from 20. This changes the relationship between beams and posts from a symmetry point of view, but this effect is dwarfed by the introduction of a dipole like asymmetry.

This design of coupler has two notable disadvantages. Similar to that in the output coupler design proposed, the dipole mode has remained ≈ 50 MHz from the operating mode with no possibility to use previously discussed dipole shifting methods (see sec. 3.4). It is however critical that any dipole element in this mode should be removed as much as possible. If a dipole-like distribution is added to the beams in the input cavity, it is possible that this will amplify down the tube leading to reduced efficiency.

Secondly a plot of E_z around a circle passing through the beam centres shows a magnitude ripple of 3.2% (see fig. 9.11). The ripple introduced by the posts can be observed but is negligible in comparison. The introduction of this variation is caused by the use of a single coupler. This could be reduced or eliminated by using a second coupler or an alternative with some kind of rotational symmetry.

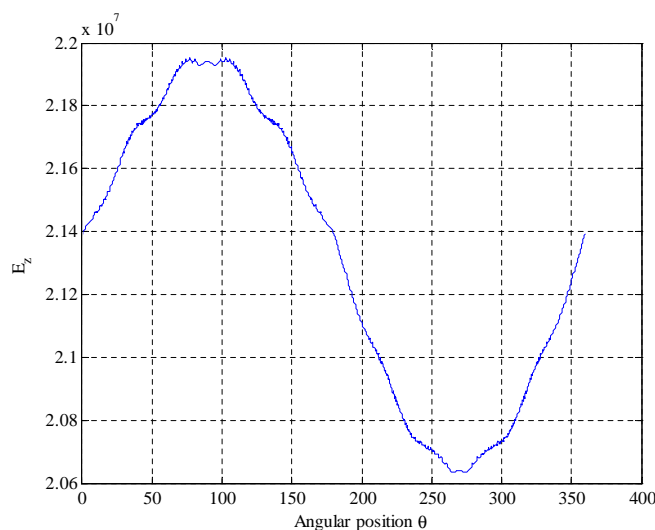


Figure 9.11: Plot of E_z around a circle passing through the beam centres. Showing the asymmetry caused by a single coupler.

Localised heating around the conducting posts could again be a problem, but the much reduced power of 200 W pulsed should reduce this heating to an easily manageable level (see fig. 9.10b).

9.4 Conclusions and Further Work

As the link between the interaction structure and the outside world the couplers are critical to a good design. In this chapter designs are proposed for both the input and output couplers.

The output coupler design is compact and allows the high power RF to leave the tube axially avoiding the solenoid. It is derived from a Calabasas Creek Research gyro-klystron coupler although in a klystron it couples to H_θ . The cavities and couplers have not been re-tuned to the target frequencies or Q s but the feasibility of the design has been demonstrated.

The output cavity easily accommodates the 15 MW design power used with a maximum theoretical power capacity of 66.353 MW, which is well above the peak power of the tube of 15 MW.

The posts used in this output coupler design cause some ripple in E_z with respect to θ (see fig. 9.6). However the input coupler has a large-dipole like ripple of 3.2% due to the use of a single coupler, which could potentially be corrected with the addition of a second coupler.

Critically, in this design, the posts act to concentrate magnetic field, particularly in the output coupler, which will lead to a high degree of localised heating. Their tight radius (mandated by the desired Q_e) renders cooling complicated.

The dipole mode remains a problem for both the input and output cavities. The previously suggested solution (see sec. 3.4) is not applicable in this case and a new approach is required, although wakefield stability analysis suggests this may not be a problem (see chap. 5). If a dipole like distribution is added to the beams in the input cavity, it is possible that this will amplify down the tube leading to reduced efficiency.

These two coupler designs appear to have merit, but have problems which must be solved before they could usefully be included in a tube.

Chapter 10

Conclusions and Further Work

10.1 Conclusions

The design of a large high power MBK remains a complex task. A number of design decisions have been made which will aid further work on this topic. In addition new techniques for rapid calculation of klystron properties have been developed which have improved the state of the art. To avoid the problems of high power design and concentrate on the issues of high efficiency operation a modified Tesla specification was used.

Table 10.1: Improved TESLA MBK parameters (values in brackets are original values)[28]

Parameter	Value	Units
Operating Frequency	1300	MHz
Klystron peak RF Power	15 (10)	MW
Beam Voltage	115	kV
Efficiency	≥ 80 (65)	%
Pulse length	1.5	msec

A target efficiency this high requires a lower perveance, multiple beams are required for this to be achieved at such a high power. This leads to a 20 beam MBK configuration, the configuration of which must be chosen.

Cavity Choice A number of cavity designs were evaluated for use as the main cavities of the MBK. As expected the re-entrant cavity did not perform well due to the large amount of excess stored energy. Refining this design to create the recessed re-entrant cavity markedly improves performance.

Two higher order mode cavities were also investigated. While they allow the radius of the tube to be increased beyond r_{pcr} with no R/Q penalty, facilitating cooling, the R/Q of the cavities is half that obtainable using a fundamental mode cavity.

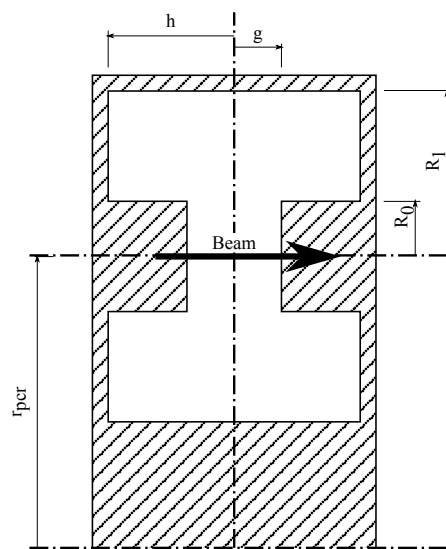


Figure 10.1: Coaxial cavity schematic (no beam pipes)

The fundamental mode coaxial cavity is the best choice of cavity from the designs investigated and is also suitable for a compatible 2^{nd} harmonic cavity. It has the advantages of the recessed re-entrant cavity in a compact, mechanically stable package. A drawback of this cavity is the proximity of the dipole mode to the operating mode. In contrast to other smaller coaxial cavity klystrons the dipole mode is only separated by 50 MHz from the operating mode.

Table 10.2: Revised parameter cavity comparison

	1.3 GHz Resonance	
	Coaxial $TM_{0,1,0}$	Coaxial 2nd Harmonic $TM_{0,1,0}$
R/Q (Ω)	130-210	>120
Q	\sim 8-9000	\sim 8000
Outside radius r_1 (mm)	160	138.5
R_1 (mm)	35-40	20
r_{pcr} (mm)	118.5	118.5
r_t (mm)	5.83	5.83
R_0 (mm)	7-21	7
Height (h) (mm)	20-27.5	15
Damping	possible	

The fundamental mode coaxial cavity can be approximately modelled as ridged waveguide. This allows a relationship between the fundamental and dipole modes to be derived using waveguide theory. This relationship explains the good performance of coaxial cavity klystrons designed by CPI to have a much smaller radius, giving them a naturally larger dipole mode separation. However, due to its size, this design will not scale to higher powers.

A similar equivalent circuit approach leads to a model which suggests that lowering the inductance (L_2) at the centre of the cavity by perturbing B_θ will increase the mode separation. This is achieved by adding a 2.5 mm slot across the centre of the cavity, which moves the dipole mode from 1.35 GHz to 1.43 GHz increasing the split from \approx 50 MHz to a much more comfortable \approx 125 MHz. The addition of lower order modes (TM_{010} , TM_{110} ...) in the centre of the cavity should not effect the operation of the klystron. As the beams do not pass through this region they should not be excited. A simple solution could be to introduce a small central post with a radius of the order of 1 mm. This would add rigidity while only perturbing these LOM. This appears to be a suitable solution to the problem of the dipole mode for the intermediate cavities. The input and output couplers may require the use of another solution.

A HOM calculation solution based on the work by Carter et al. [71]. on the method of moments was proposed in order to avoid the problems found in using *FE* or *FDTD*. This method is capable of calculating modes with no azimuthal variation to a comparable accuracy to fundamental modes. These equations were extended to include a $\cos(i\theta)$ term which allows the *TM* modes to be calculated with an accuracy of 0.01% using azimuthal variation. Modes with azimuthal variation blur the distinction between *TE* and *TM*. As a result of this the method finds modes which are in fact *TE* which break the fundamental assumptions of these equations (that $H_z = 0$). The solution to this is to augment the *TM* calculations with those for the *TE* modes, derived similarly. Combining the two sets of equations yields a solution capable of calculating many, but not all modes. Modes which are not clearly *TM* or *TE* are not represented in this set. Mixing these two modes types allows all modes to be fully represented with non zero H_z and E_z . This is achieved by altering the equations that match the fields at the interface to also mix the two sets of basis functions.

Mixed mode calculations maintain accuracy for *TM* modes with 0 azimuthal variation and increases the accuracy for an azimuthal variation of 1. The accuracy of some of these modes is however too low to be of use at almost 5%. These good results are interspersed with a number of spurious modes.

If only modes with $i = 0$ are required the method of moments is a rapid and accurate method for determining cavity properties. The solution for modes with $i \neq 0$ is more complex and can yield accurate results but they are combined with poor quality solutions.

The derivation, particularly of the coupled modes, may be simplified if performed with respect to θ as opposed to z , as the cavity has translational symmetry in the θ direction. The lack of translational symmetry in z has lead to the complexity of the equations.

Stability In order to evaluate the effect of HOM in the cavities (predominantly the dipole mode) a simple single cavity model of a multiple beam klystron was

devised. This represents the cavities as a series of independent gaps and has been constructed using standard wake-field theory. Although it ignores some complex effects it evaluates the proportion of the energy directed into the dipole mode of the cavity compared to the target operating mode.

This leads to some interesting conclusions, e.g. that lowering interaction with a mode by lowering the Q_0 is not the best solution, and can in fact increase the interaction. A much better solution is to change the frequency of the mode.

The standard 5% current error in the electron guns introduces a natural dipole component to the cavities. However this is at a very low level and even a reduction of mismatch to 1% would not eliminate the problem. The loss of a beam has the largest effect on dipole content as might be expected. The loss of the first beam is the most critical; further beam losses (even when in a group) only have a small effect.

The significance of the dipole mode content at these low levels can only be fully understood when combined with a model to take into account the gain of the klystron. A dipole mode content of -70dBc at an intermediate cavity could lead to a much larger value at the output cavity although this is unlikely. Further work could be undertaken to quantify this effect. At this stage it appears that the dipole mode is in fact not a problem.

Interaction structure An interaction structure has been designed for both high and low R/Q cavity types. It has been demonstrated that the target efficiency can be more easily achieved using a high R/Q cavity such as the fundamental coaxial cavity. The structures presented do not represent a finished design as the bandwidth, for instance, does not meet the specification. Conservative values for the input and output Q_e simplify the design of their associated couplers.

The klystron design using the initial parameters highlighted the poor choice of some parameters. This inspired a revision of these and a minor redesign. This did not adversely effect the potential performance of the klystron although it did increase its length, but the new parameters chosen greatly increase the life

of the tube.

The manual optimisation of a klystron interaction structure is a slow, time consuming task. This is a task that could potentially be automated by the use of an evolutionary algorithm. As there are a number of objectives (length and efficiency to name but two) a multi-objective algorithm must be used. This allows all of the target properties of the klystron to be optimised simultaneously.

Table 10.3: Improved TESLA klystron interaction structure

Cavity	drift (mm)	gap (2g) (mm)	F (MHz)	R/Q (Ω)	Q_0	Q_{ext}
1	0	16	1300.7	200	5000	300
2	0.6	16	1312.5	200	5000	
3	0.34	9	2544	150	3000	
4	0.508	16	1371	200	5000	
5	0.4	16	1339	200	5000	
6	0.22	16	1303	100	5000	60

Using AJDisk klystron simulation software combined with a cluster, thousands of interaction structure evaluations can be performed quickly and automatically. The techniques to enable an evolutionary algorithm to optimise a klystron, for instance Pareto Search Pressure Focusing, work well. The solution vector [efficiency, length, Δ_{gain} , u_{slow}] is used to ensure an efficient, short klystron within the specification gain, with an exit velocity above the required minimum is found.

Two klystron interaction structures have been successfully optimised, and interesting properties of the solutions space become visible. For instance the speed of the slowest electron when plotted against efficiency has a pronounced knee, showing 2 distinct regimes of output gap operation.

The Pareto fronts reveal a near linear relationship between efficiency and exit velocity in the region of interest and curves the efficiency-length and efficiency- u_{slow} plots. At this stage the convergence of the algorithm is too slow for commercial use and requires expensive computing resources. In terms of runtime it is likely to be outperformed by a skilled klystron designer, although the advantage, of course, is that their time is not required.

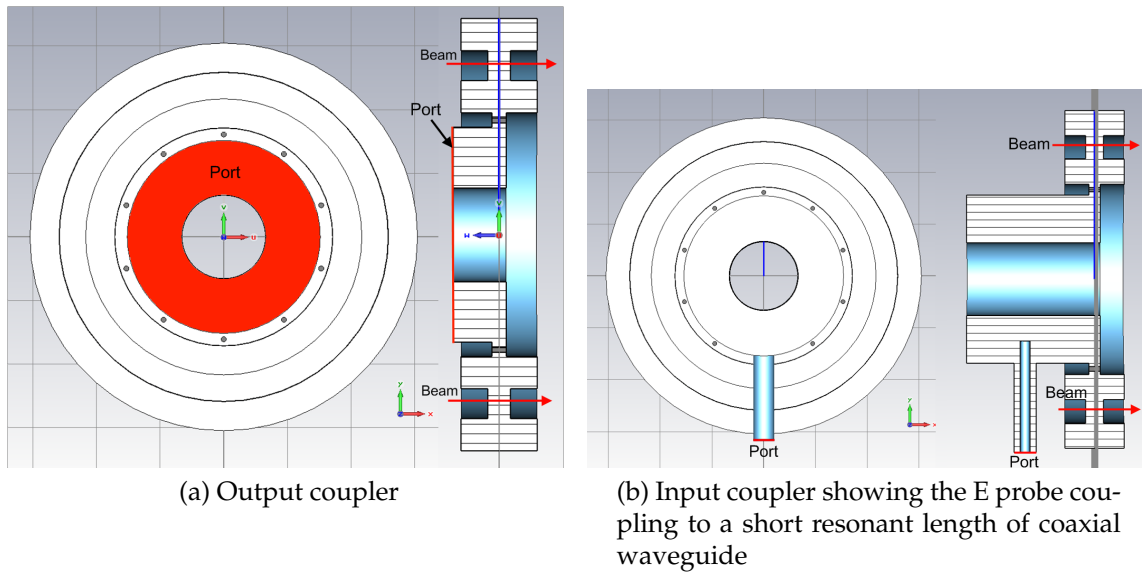


Figure 10.2: “Solid” vacuum representation of the couplers with the ports marked

Output Coupling As the link between the interaction structure and the outside world the couplers are critical to a good design. Inspired by a gyro-klystron output coupler a compact, axially exiting, MBK output coupler with the required power capacity has been proposed. While not perfectly tuned it shows that achieving the specification is possible, and demonstrates minimal ripple. The most critical flaw with this coupler is localised magnetic heating around the conducting posts which could be difficult to cool.

The input coupler uses a similar design but exhibits a 3% ripple, however the solutions may be simple. Both couplers are susceptible to dipole mode oscillations as this mode cannot be moved using the method previously proposed. This is a problem predominantly for the output cavity as a dipole element to the modulation may be introduced along the tube.

The two coupler designs appear to have merit, but their limitations must be addressed before they could usefully be included in a tube.

10.2 Further work

This thesis has not addressed the design of several major components of the klystron. The collector, gun and focusing systems for instance have not been evaluated, short of keeping their specifications modest. Their design and optimisation is a crucial step towards construction.

Although the cathode loading and spacing is conservative, design of the electron gun is not trivial. It is largely complicated by ensuring that each beam is focused and enters the klystron focusing field cleanly. This requirement complicates the design of the main focusing coil(s). Focusing pole pieces or extra focusing coils around the beams may be required to achieve a suitable trajectory, the design and evaluation of which is complex. The trajectory of the beams as they exit the focusing field also requires great care. The beam must expand sufficiently to deposit its energy evenly on the collector.

The evolutionary algorithm used to optimise the interaction structure could be applied to the cavities to automate their design. As it is multi objective the result would remain curves displaying the design choices.

Further work on the effect of HOMs on the interaction structure of an MBK, including effects such as gain through the tube, could be achieved using an appropriate code (such as TESLA or a 3D pic code such as Vorpil). This would help the HOMs of cavities to be better evaluated in the design phase. It could perhaps be proven that these modes are not significant, as the wake-field analysis suggests.

The external coupling of the tube could be further optimised to minimise ripple and the dipole component. This is of particular importance for the input cavity. A solution is required for the problem of localised heating in the the output coupler, the lack of which would lead to failure.

The innovations required for klystron optimisation, such as PSPF, would benefit from benchmarking and could prove to be a useful technique for other applications. Additional effort could also, either establish when a suitable set of

solution vectors has been reached or allow the algorithm to converge for fewer evaluations. This would markedly increase the commercial application of the technique.

Appendix A

MWS Modifications

A.1 Introduction

This section describes several extensions and techniques used in conjunction although not limited to CST Microwave Studio. They make use of the Visual Basic scripting interface to create novel features to ease the simulation load or perform new calculations

A.2 Polynomial Analysis

The initial investigation relied on brute force to optimise the cavity geometries (see sec. 2.2.1). For each free dimension a preferred range was selected, then parameter sweep was simulated over this range and the frequency and R/Q plotted (see fig. A.1). The MWS eigenmode solver includes the ability to perform a such a parameter sweep, to simulate a range of cavities with dimensions within a preset range. Using this technique maxima of certain properties and cavities with the target resonant frequency by plotting frequency and R/Q on a graph can be searched for. Unfortunately using a straight forward parameter sweep can be very wasteful as many more points are simulated than is necessary at an unnecessarily high accuracy to ensure that the R/Q has converged

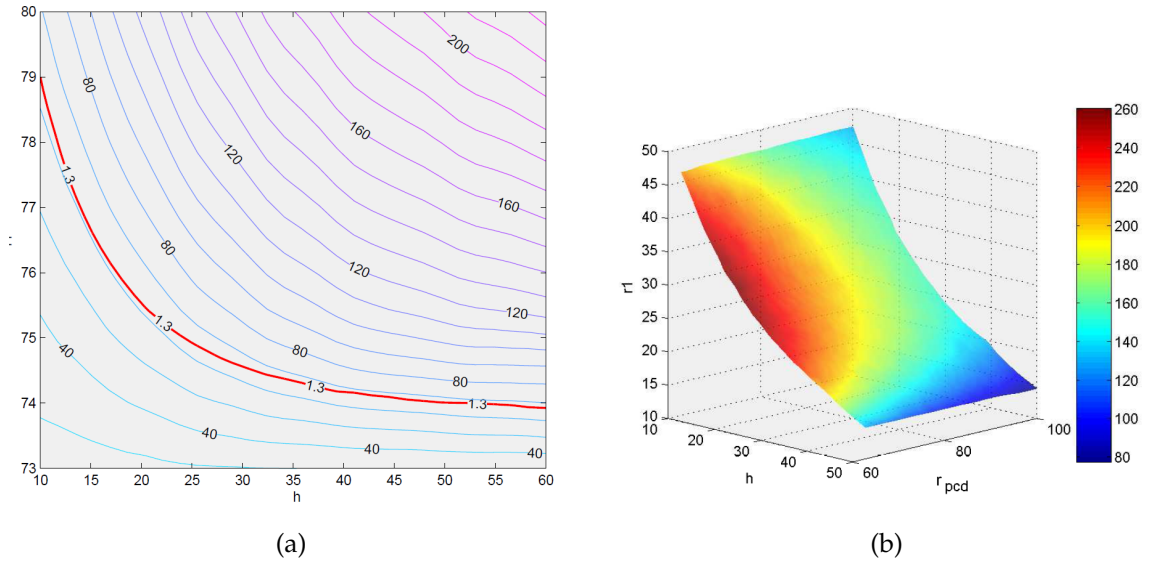


Figure A.1: The original plots which allowed selection of the optimal cavity (w.r.t. R/Q). These plots require a large number of simulations and can be hard to read.

It was noted that the density of mesh required to optimise the frequency was sparser than that required to optimise the R/Q . By plotting the results of the parameter sweeps on a contour plot of frequency against 2 other parameters and comparing to a similar graph where the contours represent R/Q , it is clear that the contours for frequency are much more reliable at a lower number of samples.

By the nature of the optimisation attempted there is only one frequency of interest. By extension the only R/Q s of interest are those of the cavities with that resonant frequency. Thus it is more economical in calculation time to take advantage of the higher reliability of the frequency contours with fewer samples and perform a much more coarse parameter sweep. Then with knowledge of the cavities of interest (those lying on the interpolated 1.3 GHz contour), calculate the R/Q s to a significantly higher accuracy.

The R/Q calculations also proved more unreliable when the revised parameters came into use (see sec. 2.9). As can be seen from the plot, the frequency contour is smooth as expected, the R/Q contours however do not allow any optimisation to be performed (see fig. A.2).

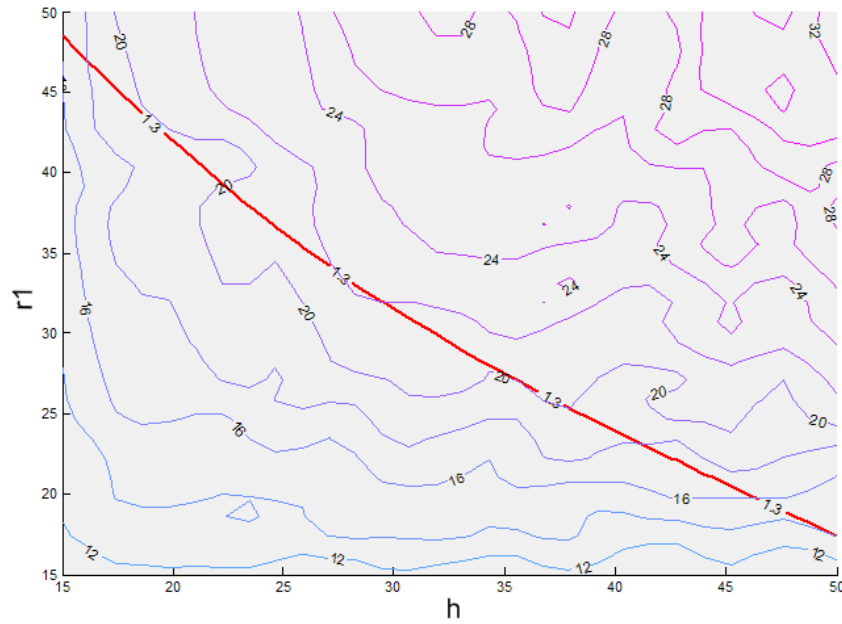


Figure A.2: Some simulation ranges can yield poorly defined R/Q curves using the brute force method (with 1.3 GHz contour). This is a clear illustration that a new optimisation method is needed.

A.2.1 Implementation - 2 Free Dimensions

MWS is used to perform the initial coarse, low mesh density parameter sweep to find the frequency variation of the cavities in the parameter space. This data set is called the "base" set. The MatLab function *griddata* performs a cubic interpolation, and *contour* to analyse the data to find the 1.3 GHz contour.

MatLab allows a contour plot to be analysed to return a series of points which lie along a particular contour, in this case the target frequency 1.3 GHz. Using these points a polynomial can be fitted using the MatLab function *polyfit*¹ which provides a formula to represent all cavities (within the constraints) which have the target frequency.

In the case of the coaxial cavity there are only 2 free dimensions with the rest being determined by the klystron parameters. This means that the result of a parameter sweep is a 2D contour plot of frequency (similar to fig. A.1

¹a 2D polynomial fitting function

(left)) although without R/Q contours. From this the contour at the frequency of interest (in this case 1.3 GHz) can be found. Using the *polyfit* algorithm in MatLab a polynomial function which relates one dimension to the other can be generated.

In order to use this function in MWS a short script using the VBA interface is needed. This script iterates over one of the variables to find the value of the other (using the polynomial approximation) which will create a cavity that resonates at 1.3 GHz. As MWS sweeps over these numbers simulating with increased mesh density it can be programmed to save the R/Q and frequency of each cavity tested. In this case two 1D plots where one free dimension is related to frequency and R/Q are created. This data set, called "line" clearly reveals the effects of changing this variable on this property of the cavity (see fig. A.4).

If frequency is plotted against the independent variable then a straight line at 1.3 GHz is expected. In reality this line is not straight due to errors introduced by the interpolation, the function fitting and the change in mesh density. The accuracy of the plot shown is within 2 MHz of the target frequency over the whole range (see fig. A.3).

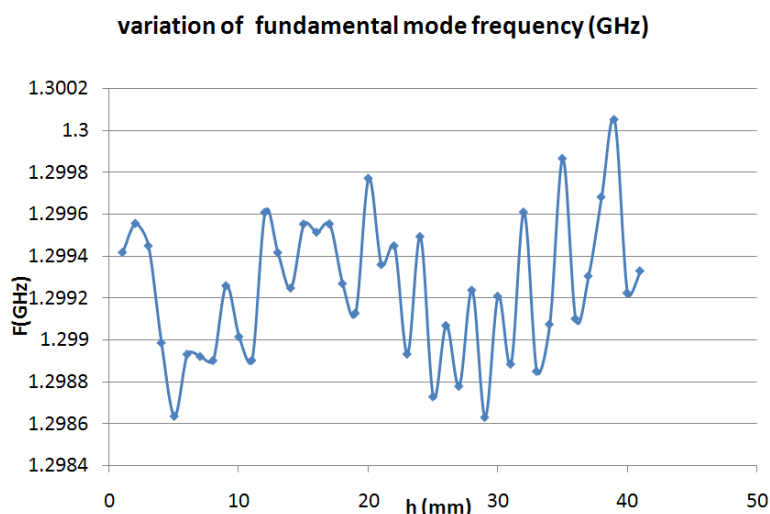


Figure A.3: Frequency stability around 1.3 GHz of a range of simulations controlled by a polynomial approximation of 1.3 GHz cavities

If the R/Q is plotted against the independent variable then the R/Q trend

can be found. The reason for the poor R/Q contours can clearly be seen from the variation of the line.

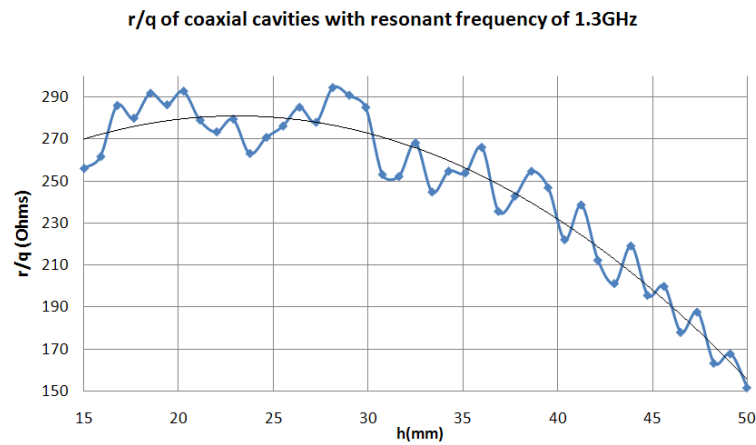


Figure A.4: Plot of R/Q against the independent variable h showing a further polynomial approximation representing the R/Q vs h . Allowing the R/Q and cavity size to be calculated from one independent variable.

A.2.2 Accuracy

The approximate accuracies of the respective optimisation methods for 2 free dimensions:

Table A.1: Simulation performance comparison

Coaxial Cavity Simulation	No. Simulations	Accuracy
Brute force	~500-1000	10 MHz
Polynomial Approximation	~50	~2 MHz

Clearly the polynomial method not only dramatically reduces the number of simulations required but it markedly increases the accuracy of the optimisation. This should of course not be confused with the accuracy between simulation and experimental measurement. This accuracy is reduced for the whispering gallery cavity because of its additional free variable.

A.2.3 R/Q Prediction

In the case of 2 free dimensions a graph of R/Q with respect to either one of the free dimensions can be plotted (see fig. A.4), a trend line can be fitted to this plot which will yield another polynomial function. This is of particular interest as it allows the prediction of the R/Q of a cavity when specifying only one of its parameters (with the other being defined by the previous polynomial function) without any further simulation.

This can also be performed in the case of the 3 free dimensions using the 3D polynomial fitting method described above.

A.2.4 Conclusions

An improved way has been found to optimise the cavities based on the fact that the frequency contours are stable with coarser grid than the R/Q contours. This method has a significantly greater frequency accuracy of results and much greater efficiency of simulation.

Appendix B

Method Of Moments

B.1 \mathbf{G}_n^{xTE} Function

$$\mathbf{G}_n^{ITE}(k) = \frac{h_n^I}{e_n^I} = \frac{j}{(\gamma_n)^2} \frac{(J_{i+1}(\gamma_n r_1) \gamma_n r_1 - i J_i(\gamma_n r_1))}{J_i(\gamma_n r_1)} \quad (\text{B.1})$$

$$\mathbf{G}_m^{II TE}(k) = \frac{h_m^{II}}{e_m^{II}} \quad (\text{B.2})$$

$$= -j \frac{-Y_i(\Gamma_2) \cdot J_{i+1}(\Gamma_1) \Gamma_1 + i Y_i(\Gamma_2) J_i(\Gamma_1) + \Gamma_1 J_i(\Gamma_2) Y_{i+1}(\Gamma_1) - i J_i(\Gamma_2) Y_i(\Gamma_1)}{r_1 \gamma_m^2 [J_i(\Gamma_1) Y_i(\Gamma_2) - J_i(\Gamma_2) Y_i(\Gamma_1)]} \quad (\text{B.3})$$

where $\Gamma_1 = \gamma_m r_1$ and $\Gamma_2 = \gamma_m r_2$.

B.2 \mathbf{H}^x and \mathbf{J}^x

The H matrices are:

$$\mathbf{H}^I = \left(\frac{i\beta^I}{r_1\gamma_n^2} \right) \text{ if } v \neq 0 \quad (\text{B.4})$$

$$\mathbf{H}^{II} = \left(\frac{i\beta^{II}}{r_1\gamma_m^2} \right) \text{ if } v \neq 0 \quad (\text{B.5})$$

The J matrices are:

$$\mathbf{J}^I = \left(\frac{i\beta^I}{r_1\gamma_n^2} \right) \text{ if } v \neq 0 \quad (\text{B.6})$$

$$\mathbf{J}^{II} = \left(\frac{i\beta^{II}}{r_1\gamma_m^2} \right) \text{ if } v \neq 0 \quad (\text{B.7})$$

B.3 Azimuthal - Variation Index (i) 2

The method of moments can also calculate modes with an azimuthal variation of 2, although its performance is significantly poorer than for $i = 1$.

B.3.1 TM Basis Functions

Table B.1: Comparison between frequencies (in GHz) from MAFIA and MoM model for azimuthal variation (i) = 2

Mode	MoM	MAFIA	Err (MWS-MoM/MoM)
1	5.67	5.85	-3.04%
2	5.92	7.38	-19.78%
3	8.75	8.74	0.05%
4	9.66	9.65	0.16%
5		9.95	$\infty\%$
6	11.6	11.6	-0.03%
7	13.0	12.8	1.97%
8	13.6	13.4	1.39%
9	13.6	13.5	0.91%
10	14.3	14.3	0.00%

Interestingly the modes which are found with an azimuthal variation of 2 show a better agreement with MAFIA 2D than with an azimuthal variation of 1.

B.3.2 Mixed TE and TM Basis Functions

The performance of the method is notably decreased for an azimuthal index of 2. Significantly modes 6 and 9 are no longer found and modes 7 and 8 are found with reduced accuracy. The accuracy of mode 1 is increased by a factor of 2 and the accuracy of modes 3, 4 and 10 is maintained.

Table B.2: Comparison between frequencies (in GHz) from MAFIA and MoM model for azimuthal variation (i) = 2

Mode	MoM	MAFIA	Err (MAFIA-MoM/MoM)	
			Mixed	TM (see table. B.1)
1	5.768873	5.85	1.45%	-3.04%
2		7.38		-19.78%
3	8.749920	8.74	0.07%	0.05%
4	9.650859	9.65	0.01%	0.16%
5		9.95		$\infty\%$
6		11.6E		-0.03%
7	11.272684	12.8E	13.35%	1.97%
8	12.60720	13.4E	6.54%	1.39%
9		13.5E		0.91%
10	14.218272	14.3E	0.51%	0.00%

Bibliography

- [1] Jean-Pierre Delahaye. The CLIC STUDY of a multi-TeV e^\pm linear collider. In *PAC99*, pages 250–252, March 1999.
- [2] M. J. Smith and G. Philips. *Power Klystrons Today*. Research Study Press Ltd., 1995.
- [3] E. Jensen, P. Pearce, and I. Syratchev. A novel idea for a CLIC 937 MHz 50MW multibeam klystron. February 2004.
- [4] George Caryotakis. Development of X-band klystron technology at SLAC. In *PAC 97*, pages 2894–2898, Vancouver, B.C., Canada, 1997.
- [5] Y. Ohkubo, H. Yonezawa, T. Shintake, H. Matsumoto, and N. Akasaka. The C-band 50MW klystron using traveling wave output structure. In *Linac 98*, August 1998.
- [6] A. Larionov, V. Teryaev, S. Matsumoto, and Y.H. Chin. Design of multi-beam klystrons in X-band. In *27th Linear Accelerator Meeting , Japan*, August 2002.
- [7] Y. H. Chin et. al. The 120MW X-band klystron development at KEK. In *EPAC 98*, pages 1894–1896, Stockholm, June 1998.
- [8] Y. H. Chin, S. Matsumoto, H. Mizuno, Y. Morozumi, T. Ohkawa, K. Ohya, K. Takata, N. Toge, S. Tokumoto, S. Kazakov, A. Larionov, and V. Teryaev. X-band PPM klystron development for JLC. In *PAC 01*, pages 3792–3794, Chicago, IL , USA, May 1991.

- [9] A. Larionov. Optical system of the powerful multiple beam L-band klystron for linear collider. In *RuPAC XIX*, pages 456–458, Dubna, October 2004.
- [10] R. Lawrence Ives, George Miram, Anatoly Krasnykh, Valentin Ivanov, David Marsden, Max Mizuhara, Tom Robinson, Ross Wilcox, and Jorge Guevara. Development of multiple beam guns for high power RF sources for accelerators and colliders. In *6th Workshop on High Energy Density and High Power RF*, volume 691, pages 90–99, Berkeley Springs, West Virginia (USA), June 2003. AIP.
- [11] CST. *CST Studio User Manual*.
- [12] T. Habermann, A. Balkcum, R. Begum, H. Bohlen, M. Cattelino, E. Cesca, L. Cox, E. Eisen, S. Forrest, D. Gajaria, T. Kimura, J. Ramirez-Aldana, A. Staprans, B. Stockwell, and L. Zitelli. Design and operation of a high power, high efficiency, l-band multiple beam klystron. *Vacuum Electronics Conference, 2009. IVEC '09. IEEE International*, pages 339–340, april 2009.
- [13] R. G. Carter. *Electromagnetic waves: Microwave components and devices*. Chapman and Hall, London, 1990.
- [14] Tobias Blicke and Lothar Thiele. A comparison of selection schemes used in evolutionary algorithms. *Evolutionary Computing*, 4(4):361–394, 1996.
- [15] Kalyanmoy Deb. A computationally efficient evolutionary algorithm for real-parameter optimization. *Evolutionary Computation*, 10(4):371–395, November 2002.
- [16] K. Gunn and C. Lingwood. The spartan evolutionary algorithm. *Technical report, Lancaster University, Engineering Department*, 2009.
- [17] K. Gunn, C.J. Taylor, and C. Lingwood. Evolutionary algorithms for the development and optimisation of wave energy converter control systems. In *EWTEC 2009*, 2009.

- [18] Kalyanmoy Deb. *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley & Sons Inc., New York, NY, USA, 2001. ISBN 047187339X.
- [19] N. Srinivas and Kalyanmoy Deb. Multiobjective optimization using non-dominated sorting in genetic algorithms. *Journal of Evolutionary Computation*, 2(3):221–248, 1994.
- [20] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, April 2002.
- [21] Eckart Zitzler, Kalyanmoy Deb, and Lothar Thiele. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary Computing*, 8(2):173–195, 2000.
- [22] E. Zitzler and L. Thiele. Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach. *IEEE Journal on Evolutionary Computing*, 3(4):257–271, November 1999.
- [23] Eckart Zitzler, Marco Laumanns, and Lothar Thiele. SPEA2: Improving the strength pareto evolutionary algorithm. Technical report, 2001.
- [24] A. Jensen. AJDISK help documentation. aaron@slac.stanford.edu.
- [25] Thales Electron Devices. Thales TH2178 Klystron Datasheet, 2009. URL http://www.thalesgroup.com/Portfolio/Security/Klystron_TH_2178/?pid=1568.
- [26] Lawrence Ives, Michael Read, Jeff Neilson, Philipp Borchard, and Max Mizuhara. 30MW 30GHz gyroklystron with an inverted MIG. In *US Workshop on High Gradient Research for Multi-TeV Linear Colliders*. Stanford Linear Accelerator Center, July 11-12, 2005.

- [27] David M. Pozar. *Microwave Engineering*. John Wiley & Sons, 2nd edition, 1997.
- [28] *International Linear Collider Technical Review Committee Second Report*. SLAC, Stanford, CA, 2003.
- [29] A. Beunas, G. Faillon, and S. Choroba. A high power long pulse high efficiency multi beam klystron. In *5th Modulator-Klystron Workshop for Future Linear Colliders*, pages e–proc, April 2001.
- [30] The CLIC Study Team. *A 3TeV $e^- e^+$ LINEAR COLLIDER BASED ON CLIC TECHNOLOGY*. Cern, July 2000.
- [31] T. K. Nelson. Precision measurements at the ILC. In *CIPANP 2006: 9th Conference on the Intersections of Particle and Nuclear Physics Reports*, volume 870, pages 232–235, Puerto Rico, 2006.
- [32] R. Brinkmann. Tesla (accelerator). In *ICFA Seminar, CERN*, October 2002.
- [33] H Braun, R Corsini, J P Delahaye, A de Roeck, S Döbert, A Ferrari, Günther Geschonke, A Grudiev, Claude Hauviller, B Jeanneret, E Jensen, T Lefèvre, Y Papaphilippou, G Riddone, L Rinolfi, W D Schlatter, Hermann Schmickler, D Schulte, I Syrathev, M Taborelli, F Tecker, R Tomás, S Weisz, and W Wuensch. Clic 2008 parameters. Technical report, CERN, Oct 2008.
- [34] R. G. Carter. Review of RF power sources for particle accelerators. In J. Miles, editor, *CERN Accelerator School : Radio Frequency Engineering*, pages 107–145, CERN, Geneva, 2005.
- [35] S. Fukuda, K. Hayashi, S. Maeda, S. Michizono, and Y. Sait. Performance of a high-power klystron using a BI cathode in the KEK electron linac. *Applied Surface Science*, 146(0):84–88, 1999.
- [36] S. Fukuda, S. Michizono, K. Nakao, Y. Saito, and S. Anami. Design and evaluation of a compact 50MW RF source of the PF linac for the KEKB

- project. *Nuclear Instruments And Methods in Physics Research*, 368:561–571, March 1995.
- [37] S. Isagawa, T. Higuchi, K. Kobayashi, S. Miyake, K. Ohya, and M. Yoshida. Application of M-type cathodes to high-power CW klystrons. *Applied Surface Science*, 146:89–96, 1999.
- [38] P.D. Pearce and A. Beunas. A klystron-modulator RF power system for the CLIC drive-beam accelerators. In *Power Modulator Symposium, 2000*, pages 161–164, June 2000.
- [39] A. N. Sandalov, V. M. Pikunov, and V. E. Rodyakin. High efficiency conventional and relativistic klystrons. In *RF 96*, pages 175–184, April 1996.
- [40] E. Wright, A. Balkcum, H. Bohlen, M. Cattelino, L. Cox, M. Cusick, E. Eisen, F. Friedlander, S. Lenci, B. Stockwell, and L. Zitelli. Development of a 10MW, L-band multiple beam klystron for TESLA. In *EPAC 2002*, pages 2337–2339, Paris, France, June 2002.
- [41] G. Caryotakis, E. Jongewaard, R. Phillips, G. Scheitrum, S. Tantawi, and N.C. Luhmann. Gigawatt multibeam klystron (GMBK) a 2-Gigawatt, 1-microsecond, microwave source. In *11th International Conference on High Power Particle Beams*, pages e–proc, Prague, Czech Republic, June 1996.
- [42] E. Jongewaard, G. Caryotakis, C. Pearson, R. M. Phillips, D. Sprehn, and A. Vlieks. The next linear collider klystron development program. In *EPAC 2000*, pages 2063–2065, Austria Center Vienna, August 2000.
- [43] George Caryotakis. The X-band klystron program at SLAC. In *17th Particle Accelerator Conference*, page 2894, Vancouver, Canada, May 1997.
- [44] D. Sprehn, G. Caryotakis, and R. M. Phillips. 150MW S-band klystron program at the Stanford Linear Accelerator Center. In *RF 96*, April 1996.

- [45] S.I. Tokumoto. High power testing results of the X-band mixed-mode RF windows for linear colliders. In *XX International Linac Conference*, pages 736–738, Monterey, California, 2000.
- [46] E. Wright, R. Callin, G. Caryotakis, K. Epply, K. Fant, R. Fowkes, S. Gold, R. Koontz, R. Miller, C. Pearson, R. Phillips, S. Tantawi, and A. Vlieks. Design of a 50MW klystron in X-band. In *Pulsed RF sources for linear colliders*, volume 337, pages 58–66, Montauk, New York (USA), October 1995. AIP.
- [47] A. Descoeur, Y. Levinsen, S. Calatroni, M. Taborelli, and W. Wuensch. Investigation of the DC vacuum breakdown mechanism. *Phys. Rev. ST Accel. Beams*, 12(9):092001, Sep 2009.
- [48] W. Peter, R. J. Faehl, A. Kadish, and L. E. Thode. Criteria for Vacuum Breakdown in RF Cavities. *IEEE Transactions on Nuclear Science*, 30:3454–+, August 1983.
- [49] S. O. Schriber. Factors limiting the operation of structures under high gradient. In *Linac 86*, 1986.
- [50] J. Rodney and M. Vaughan. Multipactor. *IEEE Transactions on Electron Devices*, 35:1172–1180, July 1988.
- [51] Gregory Nusinovich, Michael Read, and Liqun Song. Excitation of "monotron" oscillations in klystrons. *Physics of Plasmas*, 11(11):4893 – 4903, 2004.
- [52] Bruce E. Carlsten and W. Brian Haynes. Discrete monotron oscillator. *IEEE TRANSACTIONS ON PLASMA SCIENCE*, 24(4):1249–1258, August 1996.
- [53] Robert J. Barker, editor. *Modern Microwave and Millimeter-Wave Power Electronics*. Wiley-IEEE Press, 2005.
- [54] A. Yano, S. Miyake, S. Kazakov, A. Larionov, V. Teriaev, and Y. H. Chin. The Toshiba E3736 multi-beam klystron. In *LINAC 2004*, pages 706–708, Lübeck, Germany, May 2004.

- [55] A. Beunas, G. Faillon, S. Choroba, and A. Gamp. A high efficiency long pulse multi beam for the tesla linear collider. In *HEACC 2001*, page Tsukuba, Tsukuba, March 2001.
- [56] A. E. Vlieks, R. S. Callin, G. Caryotakis, K. S. Fant, W. R. Fowkes, T. G. Lee, and E. L. Wright. 100MW klystron development at SLAC. In *PAC 91*, pages 798–800, San Francisco, California, May 1991.
- [57] T. G. Lee, Gerhard T. Konrad, Yukio Okazaki, M. Watanabe, and Hiroshi Yonezawa. The design and performance of a 150MW klystron at S-band. *IEEE Transactions Plasma Science*, 13:545–552, April 1985.
- [58] T. Shintake, N. Akasaka, H. Matsumoto, Y. Ohkubo, and H. Yonezawa. Development of C-band 50MW pulse klystron for e+ e- linear collider. *IEEE Transactions on Electron Devices*, 1:533–535, 1998.
- [59] T. G. Lee. Multiple extraction cavities for high-power klystrons. *IEEE Transactions on Electron Devices*, 40:1329–1334, July 1993.
- [60] E Wright, A Balkcum, H. Bohlen, M. Cattelino, L. Cox, M. Cusick, E. Eisen, K Epply, F Freidlander, S Lenci, J Petillo, A Stamprans, B Stockwell, and L Zitelli. The VKL-8301 multiple beam klystron. In *Presented at the ITRP Visit to DESY*, 5/6th April 2004.
- [61] A. Yano, S. Miyake, K. Tetsuka, Y. Okubo, and K. Hayashi. Development of the l-band multi-beam klystron at Toshiba. (0), 0.
- [62] Simon Yu. Particle-in-cell simulation of high power klystrons. *SLAC/AP-34*, September 1984.
- [63] E. Wright, A. Balkcum, H. Bohlen, M. Cattelino, L. Cox, E. Eisen, K. Eppley, F. Friedlander, S. Lenci, A. Staprans, B. Stockwell, L. Zitelli, and K. Eppley. Test results for a 10MW, L-band multiple beam klystron for TESLA. In *EPAC 2004*, pages 1117–1119, Lucerne, Switzerland, 2004.

- [64] Hasan Padamsee. *RF Superconductivity for Accelerators*. John Wiley and Sons, INC., 1998.
- [65] Wolfgang J. R. Hoefler and Mies N. Burton. Closed-form expressions for the parameters of finned and ridged waveguides. *IEEE Transactions on Microwave Theory and Techniques*, 30(12):2190–2194, December 1982.
- [66] R.G. Carter, Jinjun Feng, and U. Becker. Calculation of the properties of re-entrant cylindrical cavity resonators. *Microwave Theory and Techniques*, 55(12):2531–2538, December 2007.
- [67] R.G. Carter, C. Lingwood, , and J.J. Feng. Rapid calculation of the properties of klystron cavities. In *2008 IEEE International Vacuum Electronics Conference*, pages 140–141, Monterey California, 2008.
- [68] E. L. Chu and W. W. Hansen. Disk loaded waveguides. *Journal of Applied Physics*, 20:280–285, March 1949.
- [69] R. Taylor. Calculation of resonant frequencies of re-entrant cylindrical electromagnetic cavities. *Journal Nuclear Energy, Plasma Physics*, 3:129–134, 1961.
- [70] Julian Schwinger and David S. Saxon. *Discontinuities in waveguides*. Gordon and Breach Science Publishers, New York, 1968.
- [71] R. G. Carter, Jin jun Feng, and Ulrich Becker. Accurate modeling of re-entrant cavity resonators. Technical report, Lancaster University, Engineering Department, 2007.
- [72] Ramo, Whinnery, and Van Duzer. *Fields and Waves in communication electronics*. Wiley, June 1965.
- [73] R. W. Hamming. *Numerical methods for scientists and engineers*. McGraw-Hill, 1962.

- [74] L. Ives, P. Ferguson, D. Marsden, Liqun Song, M. Read, G. Miram, G. Collins, A. Krasnykh, and V. Ivanov. Design, fabrication, and test of an X-band multiple beam klystron. In *Proceedings IEEE International Vacuum Electronics Conference IVEC09*, pages 341–342, April 28–30, 2009.
- [75] R. Wanzenberg. Monopole, dipole and quadrupole passbands of the TESLA 9-cell cavity. *TESLA 2001-33*, pages 9–15, September 2001.
- [76] G. Burt, R. M. Jones, and A. Dexter. Analysis of damping requirements for dipole wake-fields in rf crab cavities. *IEEE Transactions on Nuclear Science*, 54(5):1728–1734, October 2007. doi: 10.1109/TNS.2007.905787.
- [77] Wolfram Research Inc. Mathematica.
- [78] Rodople Marchesin. Meeting, 2007.
- [79] D.E Goldberg. *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley, 1989.
- [80] David E. Goldberg and Kalyanmoy Deb. *Foundations of Genetic Algorithms*. Morgan Kaufmann, 1991.
- [81] Heinz Mühlenbein and Dirk Schlierkamp-Voosen. Predictive models for the breeder genetic algorithm in continuous parameter optimization. *Evolutionary Computation*, 1:25–49, 1993.
- [82] Takahide Higuchi, Shigeyoshi Tsutsui, and Masayuki Yamamura. Theoretical analysis of simplex crossover for real-coded genetic algorithms. In *PPSN VI: Proceedings of the 6th International Conference on Parallel Problem Solving from Nature*, pages 365–374, London, UK, 2000. Springer-Verlag.
- [83] A. E. Hans. Multicriteria optimization for highly accurate systems. In W. Stadler, editor, *Multicriteria Optimization in Engineering and Sciences*, volume 19 of *Mathematical concepts and methods in science and engineering*, pages 309–352. Plenum press, New York, 1988.

- [84] James David Schaffer. *Some experiments in machine learning using vector evaluated genetic algorithms (artificial intelligence, optimization, adaptation, pattern recognition)*. PhD thesis, Vanderbilt University, Nashville, TN, USA, 1984.
- [85] T. Holland. *Persian Fire: The First World Empire, Battle for the West*. Little, Brown, 2005.
- [86] Eckart Zitzler and Lothar Thiele. An evolutionary algorithm for multi-objective optimization: The Strength Pareto Approach. Technical report, Computer Engineering and Communication Networks Lab, 1998.
- [87] Condor. High throughput computing. URL <http://www.cs.wisc.edu/condor/>.
- [88] Hidemoto Nakada. Condor JAVA API. National Institute of Advanced Industrial Science and Technology. URL http://staff.aist.go.jp/hide-nakada/condor_java_api/index.html.
- [89] Nathan Marcuvitz. *Waveguide handbook*. P. Peregrinus on behalf of the Institution of Electrical Engineers, London, UK :, 1951.