Proceedings

UNLIMITED

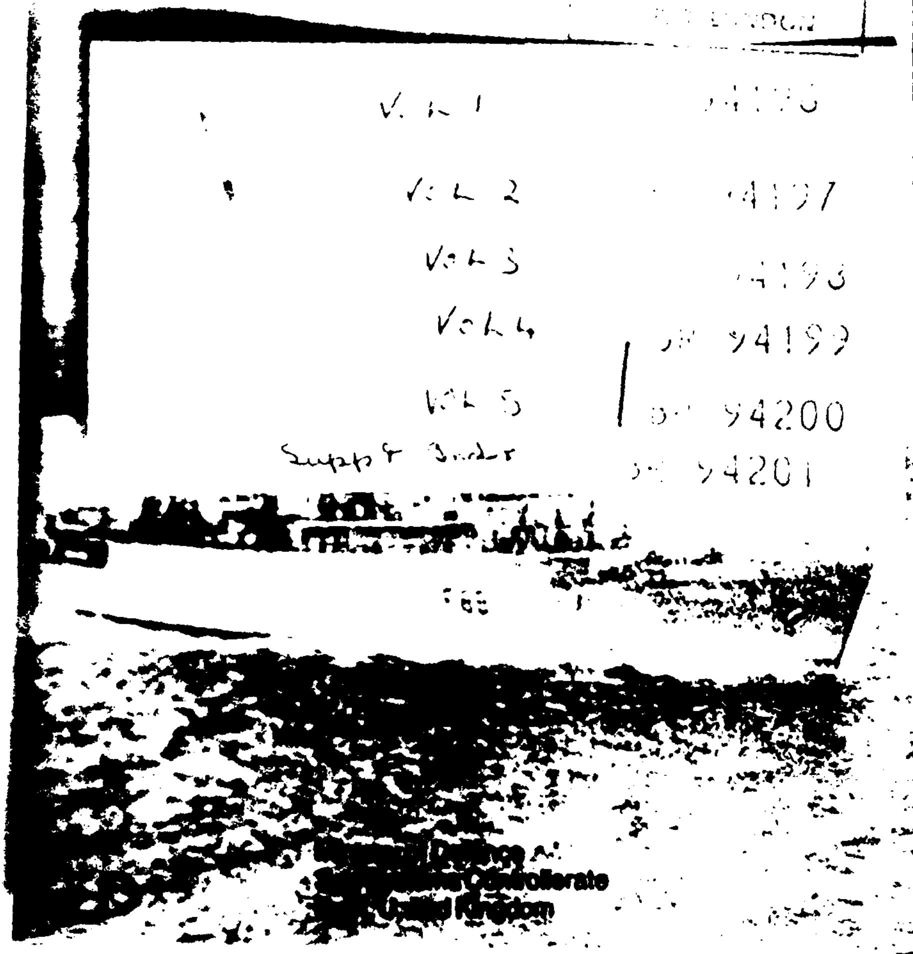
DR 94197

Seventh Ship Control Systems Symposium

24 - 27 September 1984
Bath United Kingdom

Volume 2

1984



Controllerate
United Kingdom

1984
Bath, UK

Vol 2 Proceedings of the Seventh Ship Control Systems Symposium 1984

PUBLICATION INFORMATION

These papers were printed just as received from the authors in order to ensure their availability for the Symposium.

Statements and opinions contained therein are those of the authors and are not to be construed as official or reflecting the views of the United Kingdom Ministry of Defence.

Requests for further information regarding the Proceedings and the Symposium should be addressed to:

B BLOCK, MINISTRY OF DEFENCE (PE), FOXHILL, BATH, BA1 5AB, UK:
ATTENTION: ADMC.

Volume 2

Contents

 PRECEDING PAGE BLANK

Volume 2

Contents

MULTIVARIABLE CONTROL OF A HYDROFOIL

by R Whalley
University of Bradford
and P C Gregory
Royal Naval Engineering College
Manadon, Plymouth

ABSTRACT

This paper forms a continuation study into the control of a 31 metre surface piercing hydrofoil. Using a linear stochastic model the practical consequences of adopting linear quadratic gaussian optimality theory to design a controller are discussed. A representative performance index has been obtained and the resulting system performance from the optimal approach has been compared to that obtained using a decoupling pre-compensator.

1. INTRODUCTION

Hydrofoil craft have assumed an increasingly important role as rapid transit inshore vessels. Their ability to operate at high speeds with relatively low operating costs has encouraged employment in both marine and naval roles in surveying, coastal operations and ferry work, where the larger craft can carry up to 300 passengers.

As with aircraft, hydrofoil dynamics exhibit cross coupling effects which may be exploited to achieve co-ordinated control. However, scalar classical techniques which do not attempt to integrate the movement of the various control surfaces are often used in practise in a misguided attempt to establish a simple scheme of regulation whilst disregarding operating costs.

Work by Whalley[1] et al has shown that co-ordination of the control surfaces of a modern warship can lead to considerable economies. These savings are partly as a result of reduced wear and lower maintenance and partly as a result of reductions in fuel consumption due to lower form drag. The investigation described in this paper is the second in a series of studies aimed at the implementation of such a co-ordinated control scheme on an operational hydrofoil.

A linear stochastic model obtained from sea trial data was used in the study following earlier work by Bono[2] et al wherein the identification process and model structure was discussed. In common with the control strategy employed there a similar model structure has been adopted here so that direct comparisons between the results obtained can be made.

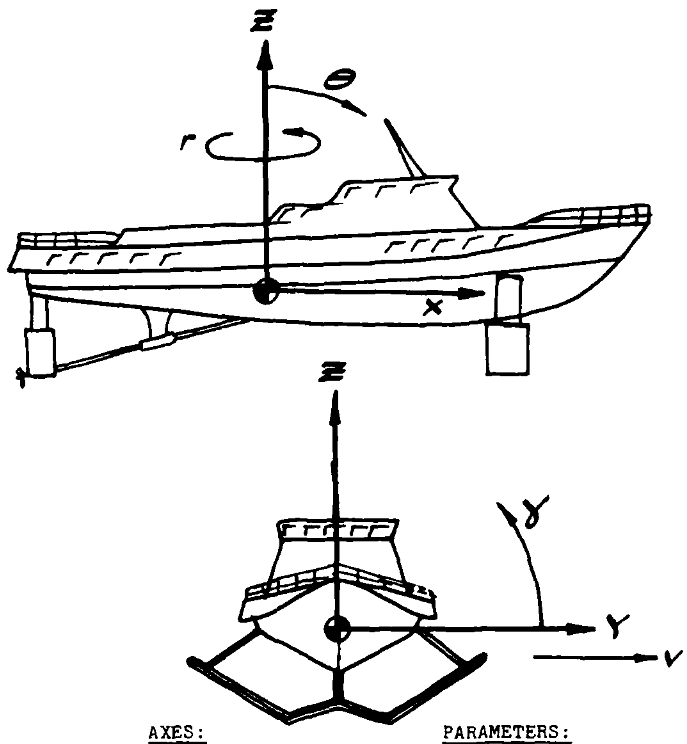
In this exposition feedback schemes based on LQG Optimal Control Theory will be investigated as part of the survey of the control strategies available and of the practical consequence of implementation.

In order to accommodate design specifications based on different performance criteria a variety of penalty functions have been provided for in the results herewith.

Comparisons between the results obtained from linear quadratic gaussian optimality theory and those obtained from simple non-interacting control theory will be made. Comments on the applicability of this form of regulation will be included.

2. MATHEMATICAL MODEL OF THE HYDROFOIL

A diagrammatic representation of a typical hydrofoil vessel is given in Figure 1 below.



AXES:

X HORIZONTAL FWD
Y HORIZONTAL PORT
Z VERTICAL UPWARDS

PARAMETERS:

θ PITCH
γ ROLL
Z HEAVE
v SWAY
r YAW RATE

Figure 1. Schematic Representation of a Hydrofoil

Linearisation of the rigid body model of the vessel gives the following state space representation.

$$\frac{d}{dt}\begin{bmatrix}\theta\\ \dot\theta\\ z\\ \dot z\\ v\\ \gamma\\ \dot\gamma\\ r\end{bmatrix} = \begin{bmatrix}0 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\ a_{21} & a_{22} & a_{23} & a_{24} & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0\\ a_{41} & a_{42} & a_{43} & a_{44} & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & a_{55} & a_{56} & a_{57} & a_{58}\\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & a_{75} & a_{76} & a_{77} & a_{78}\\ 0 & 0 & 0 & 0 & a_{85} & a_{86} & a_{87} & a_{88}\end{bmatrix}\begin{bmatrix}\theta\\ \dot\theta\\ z\\ \dot z\\ v\\ \gamma\\ \dot\gamma\\ r\end{bmatrix} + \begin{bmatrix}0 & 0 & 0 & 0 & 0\\ b_{21} & b_{22} & b_{23} & b_{24} & 0\\ 0 & 0 & 0 & 0 & 0\\ b_{41} & b_{42} & b_{43} & b_{44} & 0\\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55}\\ 0 & 0 & 0 & 0 & 0\\ b_{71} & b_{72} & b_{73} & b_{74} & b_{75}\\ b_{81} & b_{82} & b_{83} & b_{84} & b_{85}\end{bmatrix}\begin{bmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\\ \alpha_4\\ \delta\end{bmatrix} \quad .(1)$$

Sea trials on the craft performed by Istituto Per L'Automazione Navale, Genoa produced data enabling maximum likelihood methods as outlined in Kallstrom[3] to be used to estimate the coefficients:

$$a_{12} = a_{34} = a_{67} = 1$$

$$a_{21} = -1.943$$

$$a_{22} = .661$$

$$a_{43} = -.840$$

$$a_{44} = -.753$$

$$a_{77} = -.2353$$

$$a_{78} = -.534$$

and by taking the symmetry of the Hydrofoil into account

$$\begin{aligned} b_{21} &= b_{22} = -.083\\ b_{23} &= b_{24} = .096\\ b_{41} &= b_{42} = -.006\\ b_{43} &= b_{44} = .028\\ b_{71} &= -b_{72} = .063\\ b_{73} &= -b_{74} = .404 \end{aligned} \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

as detailed in Bono[2] et al.

In compliance with previous practice and since the after control surfaces have little effect on roll motion they can be ganged to provide three control surfaces.

$$u = (u_1, u_2, u_3)^t$$

where: $u_1 = \alpha_1 = \alpha_2$ = % change in after c.s. angle

$u_2 = \alpha_3$ = % change in starboard f.c.s. angle

$u_3 = \alpha_4$ = % change in port f.c.s. angle

which results in the state space description:

$$\begin{bmatrix} \dot{x} \\ \hline y \end{bmatrix} = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] \begin{bmatrix} x \\ \hline u \end{bmatrix} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

with the output equation:

$$y = Cx + Du$$

where: $$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \text{ D is null} \ldots\ldots (4)$$

and:

y_1 = % pitch angle

y_2 = % heave angle

y_3 = % roll,angle

The remainder of this paper is focused upon the construction of a feedback controller: $u = K_c(r - K_s x)$ which will induce "acceptable" transient and steady state response characteristics.

3. OPTIMAL CONTROL

The linear quadratic gaussian regulator problem is concerned with the minimisation of a performance index or functional J which comprises the weighted sum of the system state vector and the input vector. The minimisation process is subject to a constraint which is the linear system model itself, consequently:

if : $x_0 = J = \int (x^t Qx + u^t Pu)dt$

and : $\dot{x}_j = f_j(x, u, t) \quad 1 < j < n$

then : $\dot{x}_k = f_k(x, u, t) \quad 0 < k < n$

where : $\dot{x}_0 = x^t Qx + u^t Pu$

$f_j = Ax + Bu \quad 1 < j < n$

The solution to this problem is therefore a special case of the Hamilton-Jacobi equation which as shown in Athens and Falb(4) leads to:

$$\dot{R} + Q - RBP^{-1}B^TR + RA + A^TR = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots - (5)$$

A steady-state solution to this equation is given by Potter[5] establishing the control law:

$$u = -P^{-1}B^TRx \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \ (6)$$

4. CHOICE OF A PERFORMANCE INDEX

There does not appear to be any rational approach to selecting performance indices for control problems of the type considered here. For example Reid[6] uses a performance index for a surface ship steering scheme of:

$$J = \lim_{T \to \infty} \frac{1}{2T} \int_0^T \begin{bmatrix} \psi & \delta \end{bmatrix} \begin{bmatrix} \lambda & \\ & 1 \end{bmatrix} \begin{bmatrix} \psi \\ \delta \end{bmatrix} dt$$

where λ a weighting coefficient penalising yaw deviations, varies from 29.5 to 3.9 depending upon full load or ballast conditions whereas Cuong[7] used a performance index of:

$$J = \tfrac{1}{2} \int_{t_0}^{t_f} (x^TQx + u^TPu)dt$$

with $q_{44} = 772.5$, $q_{55} = 131.3$ (all other terms zero) and $p_{11} = 131.3$ for a similar scheme on a similar vessel.

Grimble[9 & 10] on the other hand uses a performance index of similar form to Cuong which, in the case of the vessel Star Hercules, has $P = I_2$ and Q a 6×6 matrix with $q_{22} = 1.759$, $q_{44} = .0487$ and $q_{24} = q_{42} = -.1755$ and in the case of the oil rig drilling vessel Wimpey Sealab has $p_{11} = p_{22} = 2000$, $p_{12} = p_{21} = 500$ and a low frequency solution for Q given as $q_{11} = 100$, $q_{22} = q_{44} = 200$, $q_{33} = q_{55} = q_{66} = 200$.

To avoid the issue a variety of performance indices have been examined. Weighting ranged from:

$$\begin{array}{lll} 10 \to .0025 & \text{for } P & \text{diagonal} \\ 800 \to -800 & \text{for } Q & \text{diagonal} \end{array}$$

For purposes of comparison with the simple de-coupling strategy proposed in Bono[2] et al a coincidental Eigenvalue set was selected corresponding to a performance index of

$$J = \int x^T \begin{bmatrix} 97 & 0 & 0 & 0 & 0 & 0 \\ 0 & -38.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 420 & 0 & 0 & 0 \\ 0 & 0 & 0 & -400 & 0 & 0 \\ 0 & 0 & 0 & 0 & 17 & 0 \\ 0 & 0 & 0 & 0 & 0 & -6 \end{bmatrix} x + u^T \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} u \ dt$$

The solution of the matrix Riccati equation given at equation 5 for this performance index and the state space model given in equation 3 establishes the control law of equation 6 where:

$$u = \left[-P^{-1}\Delta^T R\right]x = \begin{bmatrix} -3.35 & -.112 & -.619 & .855 & 0 & 0 \\ 1.98 & -.0838 & 4.99 & .0569 & 1.21 & -.00637 \\ 1.98 & -.0838 & 4.99 & .0569 & -1.21 & .00637 \end{bmatrix} x$$

5. DIVIDED GAIN METHODS

In Whalley[8] it is shown that provided the optimal gain matrix is distributed between the error channel and the feedback path in accordance with the equations:

$$K_e = -\left(C(A+F)^{-1}B\right)^{-1}$$

$$\text{where} \quad F = -BKC$$

$$\text{and} \quad K = K_e . K_f .$$

then the steady-state coupling following step changes on any input will be zero. Moreover since the loop gain remains constant at K the stability margins of the system remain unaltered.

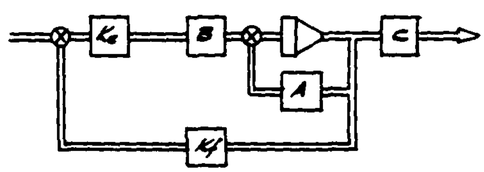
In accordance with these formulae the gain division for zero steady-state interaction whilst minimising:

$$J = \int \left(x^T[Q]x + u^T[P]u\right)dt$$

guarantees the direct relationship between set points of corresponding outputs under quiescent conditions.

6. COMPARATIVE STUDY

In block form the optimal control scheme is as shown in Figure 2. With error channel and feedback path gains given as in equations 7 and 8.

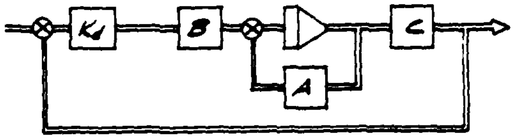


$$K_e = \begin{bmatrix} -6.48 & 4.31 & 0 \\ -.407 & 4.79 & 1.24 \\ -.409 & 4.79 & -1.24 \end{bmatrix} \qquad (7)$$

$$K_f = \begin{bmatrix} .864 & [illegible] & .85 & -.134 & 0 & 0 \\ .496 & -.0169 & 1.18 & -.00587 & 0 & 0 \\ 0 & 0 & 0 & 0 & .978 & -.0085 \end{bmatrix} \qquad (8)$$

Figure 2. Optimal Control Scheme

Whereas the non-optimal scheme discussed in Bono[2] et al is given in Figure 3 with the single gain structure given in equation 9. Equation 9 also guarantees steady-state de-coupling which gives direct correspondence between the step inputs and outputs.



$$K_d = \begin{bmatrix} -8 & 3.5 & 0 \\ -1.72 & 4.75 & 1.238 \\ -1.72 & 4.75 & -1.238 \end{bmatrix} \qquad (9)$$

Figure 3. Non-Optimal Control Scheme

For purposes of comparison the transient responses following step inputs on each reference set point in turn are shown side by side for the optimal and non-optimal schemes in Figures 4, 5 and 6.

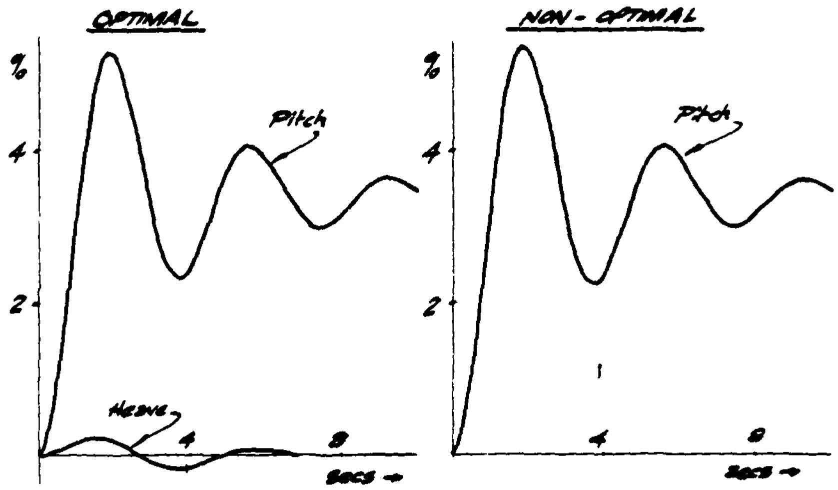


Figure 4 OPTIMAL & NON-OPTIMAL SYSTEM RESPONSE TO 10% STEP INPUT TO AFT CONTROL SURFACES

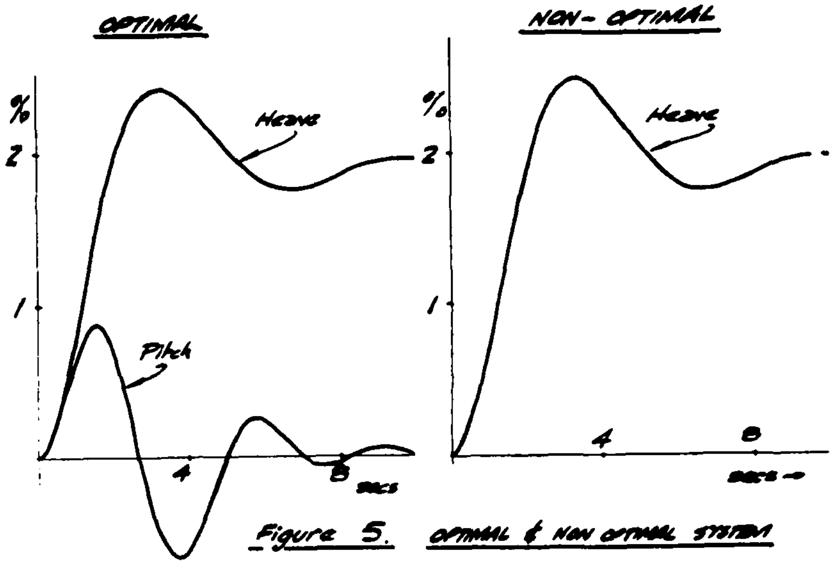


Figure 5. OPTIMAL & NON OPTIMAL SYSTEM RESPONSE TO 10% STEP INPUT TO STBD FWD CONTROL SURFACE

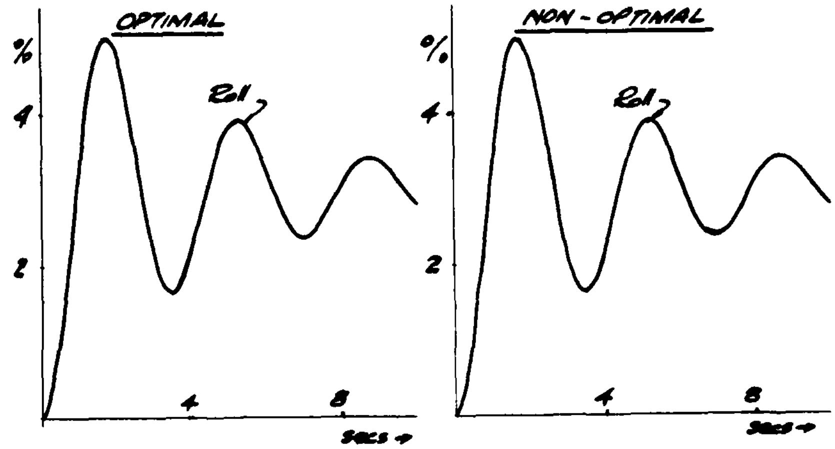


Figure 6. OPTIMAL & NON-OPTIMAL SYSTEM RESPONSE TO 10% STEP INPUT TO PORT FWD CONTROL SURFACE

These traces highlight the distinct disadvantage in employing an optimal controller in that cross coupling is present, particularly in response to a step change in the port forward control surface which produces a 45% transient interaction.

7. CONCLUSIONS

An optimal control scheme has been designed which has the same closed loop poles and the same steady-state response following a step input, as that obtained when using a simple pre-compensator.

In order to obtain a comparative performance index it was necessary to evaluate the effect of altering performance index matrices P and Q by producing an Eigenvalue migration plot and thus match 'optimal' closed loop poles to 'decoupled' closed loop poles.

The optimal control scheme is complicated, it does not wholly eliminate the high degree of natural cross coupling evident in the system as shown in Bono et al (2) and transient interaction of up to 45% occurs in the case of a step input to the port forward control surface. Transient performance characteristics in both the optimal and non-optimal configurations were similar, as shown in Figures 4, 5 and 6, as expected, but the Performance Index means little in practical terms, nor does it have any correspondence with the figures used in the other papers referenced.

The problems of determining a meaningful Performance Index are clearly considerable and Filters or Observers to estimate those of the six state variables which cannot be directly measured would be required if the Optimal approach were considered. Conversely the more practical frequency response methods now available enable simpler feedback schemes to be employed to greater effect using the measured outputs, in this case of pitch, roll and heave, which are of direct interest.

Finally it should be noted that the problems of noise and integrity have not been addressed in this paper. McMORRAN(11) et al illustrates that the failure of one state measuring instrument may cause the system to become unstable whereas graceful degradation is guaranteed by the Inverse Nyquist Array method, for example. There appears therefore to be no advantage whatsoever in using an optimal control strategy for this application.

8. REFERENCES

1. WHALLEY R, FREEMAN D J, WAUGH M G, Ship Motion and Computer Control, Proc 4th IFAC Symposium Ship Operation Automation, Genoa, Italy, Sep 1982.

2. BONO R, FUZZARD R L, TIANO A, WHALLEY R, Multivariable Identification and Control of a Hydrofoil, IASTED Symposium, Copenhagen, 1983.

3. KALLSTROM C G, ISSEBO T, ASTROM K (1976), A Computer Program for Maximum Likelihood Identification of Linear, Multivariable Stochastic Systems, Proc. 4th IFAC Symposium on Identification and System Parameter Estimation, Tblisis, USSR.

4. ATHENS M, FALB P L, Optimal Control, McGraw-Hill Book Co, New York 1966.

5. POTTER J E, Matrix Quadratic Solutions, J SIAM Applied Mathematics, Vol 14 No 3, May 1966.

6. REID R E, MEARS B C, WISE K A, TUGCU A K, GRIFFIN D E Comparison of Automatic Steering Performance of a VLCC in a seaway resulting from Application of LQG and Classical Control System Design Techniques, Proc 6th Ship Control Systems Symposium, Ottawa, Canada, 1981.

7. CUONG H T, PARSONS M G, Surface Ship Path Control using Multivariable Integral Control, Proc 6th Ship Control Systems Symposium, Ottawa, Canada, 1981.

8. WHALLEY R, An Analytical Technique for Multivariable Regulator Design. Control Section Report 33. RNEC, Manadon, Plymouth, Jun 1976.

9. GRIMBLE M J, Structure of Large Stochastic Optimal and Sub-Optimal Systems, IEE Proc. Vol 129, Pt D, No 5, Sep 82.

10. GRIMBLE M J, PATTON R J, WISE D A, The Design of Dynamic Ship Positioning Control Systems using Stochastic Optimal Control Theory, Optimal Control Applications & Methods, Vol 1, 1980.

11. ROSENBROCK H H, McMORRAN P D, Good, Bad or Optimal, IEEE Transactions on Automatic Control, Vol AC-16, No 6, 1971.

VERTICAL AND HORIZONTAL PLANE CONTROL OF SWATH SHIPS

by
James A. Fein
David W. Taylor Naval Ship R & D Center
Bethesda, Maryland, USA

ABSTRACT

Small Waterplane Area Twin-Hull (SWATH) ships offer advantages over conventional monohulls in terms of improved motions in a seaway and increased deck area without sacrificing other performance characteristics. The size and location of control devices such as rudders and fins can have a strong impact on the overall viability of the design. The control system must be carefully designed in order to assure that the full potential of the SWATH concept for low motions and adequate maneuverability is realized. This paper addresses maneuverability and motion control through the use of active control fins. Experimental and analytical information is presented for rudder configurations and rudder positions. The use of rudders behind the propeller in conjunction with an extended strut, and the use of differential thrust are discussed. Results are given for turning by U.S., Canadian, and Japanese designs that show maneuvering capabilities equivalent to conventional monohulls. The paper addresses a number of motion control issues including fin size and location and variation of control strategy. The reductions in motions when automatic fin control is applied to the SWATH are demonstrated.

WHAT IS A SWATH?

The SWATH ship is a displacement type hull form that departs significantly from conventional ships yet draws on the technology and design experience of traditional naval architecture. It offers high performance by providing platform steadiness and sustained speed capability in waves.

An Englishman, Creed, first brought the SWATH concept to the attention of the world in 1943, as a superior platform for aircraft operations. U. S. interest began in the late sixties with the design and model testing of the TRISEC by Leopold (1) of Litton Industries and the design and construction of a SWATH workboat, the SSP KAIMALINO (2). In the last few years, interest in SWATH has grown worldwide with major efforts by Mitsui Shipbuilding in Japan, the U. S. Coast Guard, the Canadian Navy, the British Navy and numerous universities and research establishments.

The acronym SWATH (Small Waterplane Area Twin-Hull) was selected in 1972 to reduce confusion between this concept and conventional catamarans, which are different in many important respects. Physically, the most apparent differences between the two concepts are below the waterline. Conventional catamarans have more or less standard displacement ship hulls, while each SWATH lower hull consists of a submerged cylindrical body connected to one or two slender surface piercing struts. The SWATH ship possesses a large deck area by virtue of the twin hull configuration. This deck serves many useful civilian and military purposes. The SWATH also has inherently low motions in waves. Ship motions are forced oscillations excited by forces generated by the waves the ship encounters. Motions vary with the geometry of the ship, particularly the distribution as well as amount of waterplane area (horizontal cross-sectional area at the waterline) and waterplane inertia in relation to ship mass and draft. Generally, less waterplane area for a given displacement will result in less wave excitation force and longer ship natural periods. Indeed, with proper design, heave and pitch wave excitation forces can be reduced to nearly zero over a narrow range of wave encounter frequencies.

The configuration of a SWATH ship is fundamentally a streamlined adaptation of a column-stabilized platform, though the SWATH ship is smaller and has relatively greater waterplane area. The SWATH geometry results in natural periods longer than those of most ocean waves in moderate storm conditions, but shorter than the periods of many waves in severe storms. However, compared to conventional ships, the SWATH configuration extends considerably the range of wave conditions and ship speeds in which excellent seakeeping qualities can be maintained. In addition, the operator of a SWATH ship will have much less eed to change heading because of excessive rolling or deck wetness. It is now possible to tailor the motion characteristics of particular SWATH designs to expected operating environments and mission speeds. This more sophisticated approach will result in a ship of enhanced operability, i.e., increased probability that the ship and all essential equipment can function properly to carry out the intended mission in adverse conditions.

SWATH MANEUVERING

Maneuvering is an issue for SWATH ships, while it is often taken for granted by conventional monohull designers. The SWATH as a catamaran has widely spaced propulsion, encouraging the use of differential thrust for superior maneuverability at very low speeds. At high speeds, the distribution of strut area can lead to a high degree of directional stability since the centroid of the strut is usually aft of the CG of the ship as a whole. Thus it takes a substantial force to generate the side force required to initiate a turn at high speed. The high degree of directional stability is an advantage for missions that require holding oblique headings to waves.

The use of active fins to reduce the motions of SWATH in a seaway provides an indirect benefit for the maneuverability at high speed in that the fins can induce roll into the direction of the turn. This improves the turning performance by increasing the drag on the side inboard to the turn which provides yawing moment. The magnitude of the induced roll is limited by the broaching of the lower hull on the outside of the turn.

The predictions for the turning of early SWATH models led to the feeling that the concept had poor maneuverability. Turn diameters given in ship lengths were misleading since a SWATH ship is likely to be shorter than a conventional ship of the same displacement or designed for the same mission. Early SWATH model designs, configured for minimum drag, had very small and ineffective rudders.

The SWATH ship is capable of turning performance equal to that of conventional monohulls designed for the same mission and speed. However the SWATH designer has to choose between a number of rudder configurations and maneuvering approaches in order to assure that acceptable maneuvering will be achieved. First the rudder configurations will be described, then the results for full scale trials, free running models, and maneuvering analyses will be given.

Rudder Configurations

The rudder design issue for SWATH ships has presented problems from the first designs to the present time. The TRISEC used an X-tail located on the lower hull just in front of the propeller, similar to some submarine control configurations. This meant that control deflections induced both horizontal and vertical force components, leading to the need for complex strategies to assure adequate turning without unacceptable vertical motions. The X-tail fins were limited in size since otherwise they would extend below the keel line of the lower hulls and outboard as well, which would lead to potential problems in docking and operating in shallow water. Another concern with this configuration was that fins mounted close to the propellers and close to the free surface could be susceptible to ventilation and could adversely affect the flow into the propeller. These considerations have led to the result that neither X-tails nor fins that extend either outboard or below the keel line have been utilized in designs since the TRISEC, with the exception of the retractable turning foil to be discussed subsequently.

The SSP KAIMALINO utilized a rectangular rudder mounted in the propeller slipstream. This configuration has the advantage that the flow induced by the propeller increased the effectiveness of the rudder. This added effectiveness is important at low speeds where rudders outside the slipstream might be ineffective. The configuration requires an "extended strut" to support the rudder. This addition to the strut can add to the cost of the platform. The strut extension must be structurally capable of of supporting the forces generated by the rudder and must house the steering gear or shafting. The strut extension may also degrade the flow into the propeller, and will increase the overall drag of the ship. The main advantage to the extended strut configuration with rudder behind the propeller is that a relatively small rudder can assure adequate maneuverability and directional control. This configuration is also the most like a conventional monohull ship steering arrangement. Mitsui Shipbuilding of Japan which has built the largest SWATH ship, the SEAGULL, and is currently constructing an even larger vessel, has utilized the extended strut with rudder mounted behind the propeller for its designs.

Related to the consideration of extended struts is the alternative of single vs twin struts. This refers to the numbers of struts on each side of the SWATH, i.e., the introduction of a gap in each side as opposed to a continuous strut on each side. The gap between the struts may mitigate motions in beam seas at very low speeds. But at higher speeds the twin strut configuration tends to increase the wave making drag significantly, though wetted surface is lower than for a single strut per side. From a maneuvering standpoint, the choice of strut number and location affects the inherent directional stability of the design. A twin strut design lends itself to the use of extended strut to place the rudder behind the propeller as in the case of the SSP KAIMALINO. However, drag considerations usually take precedence over other factors in the choice of number of struts. SWATH ships built since the SSP KAIMALINO have all utilized the single strut. There is no inherent advantage to either number of struts from a maneuvering standpoint. It is distribution of strut area with respect to the ship CG that dictates directional stability and it is size and location of the rudder that influences the turn performance.

A number of rudder configurations have been considered for SWATH ships. Those illustrated in Figure 1 were the subject of captive model tests. The spade rudder, strut rudder, and turning foil were the subject of extensive model scale evaluations utilizing a rotating arm facility (3). The spade rudder aft of the propeller has been the subject of a recent

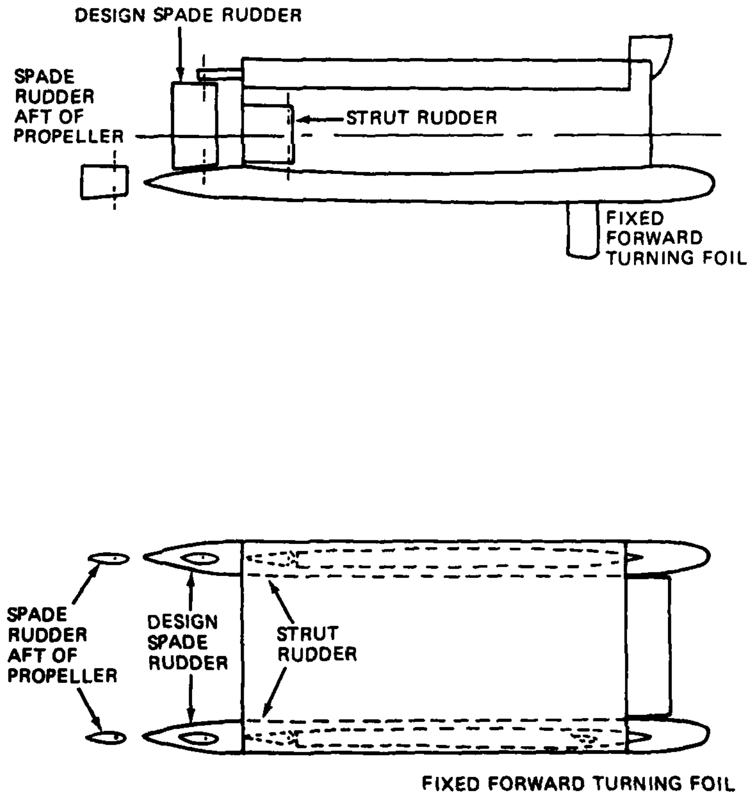


Figure 1. Schematic of SWATH-6 Model with Rudder Configurations.

model scale evaluation (4) utilizing an extended strut modification to the SWATH 6 model. It was found to have superior turning performance over the whole speed range when compared to rudders of equal area. This emphasizes the advantages of the extended strut approach, although due to the structural implications of hanging a rudder behind a propeller on a SWATH, the equal rudder area comparison may be questionable.

The strut rudder was an attempt to provide turning with minimum adverse impact on the drag of the ship. For straight ahead operation the configuration is hydrodynamically similar to a rudderless SWATH. The strut rudder, however, did not prove to be effective at either high or low speeds. The size of the rudder required for turning was a problem as was the force required to turn it. By its nature as a "flap" on the strut, it could not be balanced, thus leading to high levels of rudder torque. The strut rudder had the advantages of minimum drag, minimum cost of construction and simplicity, but could not provide adequate turning particularly at high speed.

Two types of spade rudders were studied, the surface piercing one given in Figure 1 and a much smaller fully submerged spade rudder. The fully submerged spade rudder was too small to be truly effective and its configuration would have required steering gear in the lower hull. The surface piercing spade rudder did provide adequate turning force and allowed the steering gear to be placed in the upper hull above the rudder. At high speeds, the surface piercing spade rudder tended to ventilate at large rudder angles (above 25 degrees). This caused large periodic force pulses to impinge on the propeller, leading to possibly unacceptable loads on the propulsion system. Means of alleviating these periodic forces have not been found other than restricting the maximum rudder angle at high speeds. Another problem with the spade rudders was that at some speeds the free surface behind the strut dipped, diminishing rudder effectiveness.

The forward retractable turning foil is essentially an augmenting device that is particularly compatible with the spade rudder configuration. It consists of a cambered foil in the forward section of each hull that can be employed on the inboard side to initiate a high speed turn. The foil does not deflect but the exposed area can be varied. The foil produces turning moment by both drag and side forces to yaw the ship into the turn. The concept has shown the ability to reduce turn circles by 20 percent. Its main drawbacks are cost and the possibility that the foil would become stuck, increasing the draft of the ship to an unacceptable level.

Another approach to maneuvering is to use the vertical plane control surfaces canted at some angle to induce turning. This technique has not been experimentally verified but has similarities to the TRISEC approach. It offers the potential to eliminate the rudder but may degrade the control of vertical motions since some control authority must be reserved for turning. This approach would require a combined maneuvering, coursekeeping, and motion automatic control system. Other ideas include the use of bow thrusters, flaps at the forward end of the struts or combinations of various approaches. No clear cut answer is available, thus rudder configurations for SWATH ships will depend on mission requirements and will offer the designer more choices than are available for conventional ships.

Experimental Maneuvering Results

Up to this time SWATH maneuvering results have been experimentally based, either directly by full scale or radio-controlled model results or indirectly by simulations based on empirically derived coefficients.

Figure 2 contains a summary of tactical diameter results normalized by length from three direct sources - the SSP KAIMALINO trials, the Mitsui SEAGULL trials documented by Narita (5), and Canadian radio-controlled model experiments (Reference (6)) on two rudder configurations for a SWATH 6 model similar to that in Figure 1. Figure 3 shows the corresponding speed loss ratio for the turns given in Figure 2. The figures show results for the Canadian 6A configuration which has a spade rudder configuration much like that shown in Figure 1 and the Canadian 6E which is an extended strut version of the small model with rudder behind the propeller. The ratio is the speed loss divided by the initial speed prior to the turn. The 6A has much larger tactical diameters at the higher speeds and shows a lower speed loss throughout the speed range. The SSP KAIMALINO with a rudder behind the propeller has a tactical diameter that is uniformly greater than the 6E over the overlapping Froude Number range. This difference is attributable to differences in geometry such as length, beam, rudder area and other factors. The 6E tactical diameter-length ratio of five in its operating Froude Number range is indicative of superior turning performance. A conventional monohull combatant typically would have a tactical diameter of seven ship lengths at high speed. The tactical diameter results for the Mitsui SEAGULL are interesting. It is the largest SWATH ship built at 670 metric tons. The SEAGULL also has a rudder behind each propeller supported by an extended strut. Geometrically it is similar to the 6E and the results seem to follow the trends of the 6E although the Froude Number ranges do not overlap.

The maneuvering prediction computer program summarized in Reference (3) has been utilized to investigate and compare the various rudder configurations for SWATH. The program is limited by the extent of the data base, but is adequate for comparative purposes. A comparison was made of the maneuvering performance for a 3000 metric ton single strut design with either the surface piercing spade rudder or the rudder aft-of-propeller. The tactical diameter using the aft rudder location is 30 percent less than that with the spade rudder. The speed loss for the rudder "aft" is larger for a given rudder angle; however, that rudder configuration can achieve a given tactical diameter with a smaller rudder deflection and at a higher speed. A comparison of the performance of the two rudder configurations for various speeds is presented in Figure 4 for simulated turns from initial speeds of 10, 20, and 30 knots. The "aft" configuration shows better performance over the entire speed range, particularly at higher speeds. This configuration does not show as large an increase in tactical diameter with speed as the baseline spade rudder. At low speeds the propeller slipstream improves the performance of the rudder "aft" configuration by maintaining a high velocity over the control surface and therefore increasing the control force while the hydrodynamic hull forces drop due to the slower flow past the hull.

To compare the maneuvering performance of the single strut and the twin strut concepts, the twin strut data base was scaled up to simulate a 3000 metric ton SSP. The geosym was Froude scaled. The single strut concept was modeled with a standard spade rudder between the strut and the propeller, while the twin strut concept was modeled using the rudder aft-of-propeller as is present on the SSP. Figure 5 presents the results of the comparison of the two concepts over a range of speeds for a rudder deflection angle of 25 degrees. The variation of tactical diameter with speed for the two designs indicates that for a reasonable forward speed the SSP has comparable maneuvering performance. Which strut configuration offers the best potential for turning performance is a more complex issue. It depends more on distribution of the strut area fore and aft of the CG than on the number of struts.

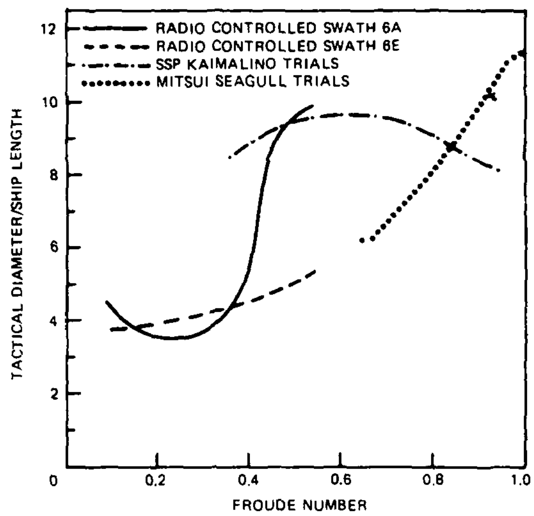


Figure 2. SWATH Tactical Diameter Results.

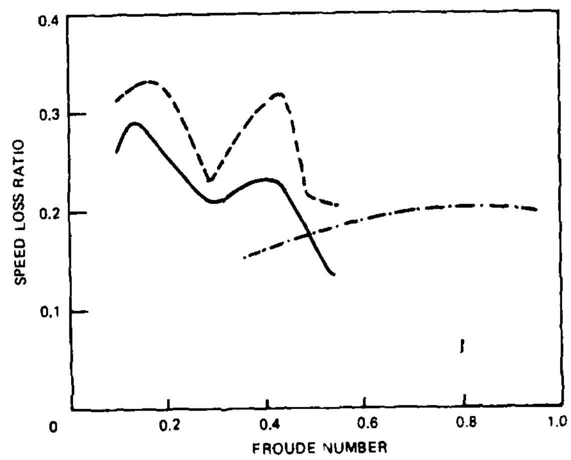


Figure 3. Speed Loss in a Turn as a Percentage of Initial Speed.

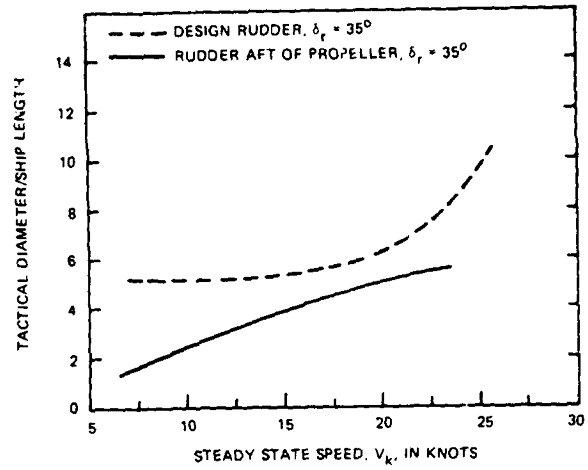


Figure 4. Predicted Tactical Diameters for the SWATH 6 as a Function of Rudder Position.

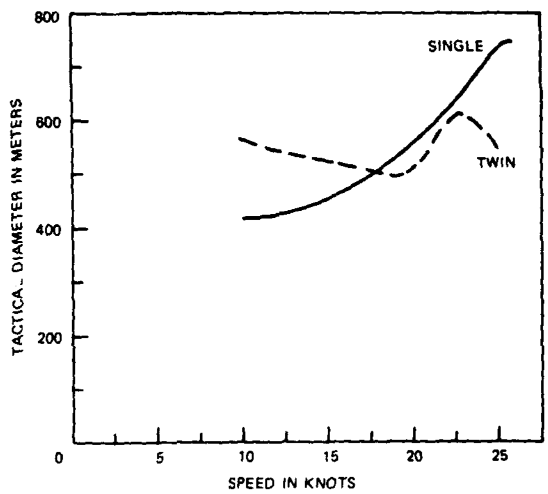


Figure 5. Predicted Tactical Diameters for 3000 metric ton Designs as a Function of Struts Per Side.

Low Speed Maneuvering

At low speeds the preferred control strategy is to use differential propeller thrust to turn the vessel. SSP KAIMALINO trials results (3) show turns within the ship's own length at very low speeds. At speeds below 10 knots, there is usually a large amount of unused propeller thrust available for control. If good low speed maneuvering performance is required, then it is necessary to be able to apply a large amount of the thrust available quickly through the use of controllable pitch propellers. Use of variable RPM with fixed pitch propellers for low speed turning is not advisable due to the loads this would create on a conventional gearbox. Variable RPM from an electric transmission system, on the other hand, shows great promise.

Bow tunnel thrusters could be used to augment the turning available from the propellers. Such a combined system tied to automatic control would have exceptional stationkeeping and docking capabilities even in high sea states.

SWATH MOTION CONTROL

The SWATH ship possesses good motion characteristics for a number of reasons. Resonant conditions, when the encountered wave period matches the ship's natural period, occur rarely. The force generated on the hull by the waves is low due to the small waterplane area. This is similar to the case with column-stabilized drilling platforms. The SWATH ship, however, since it is capable of sustained forward speed, also benefits from the pitch and heave damping generated by struts, the lower hulls, and control surfaces. Thus, the SWATH designer has a number of options available to reduce motions in waves.

Control Surfaces

All SWATH designs in recent times have utilized four control surfaces, mounted inboard, one forward and one aft on each hull. The initial purpose of the control surfaces was to assure dynamic stability at speed. Very early SWATH designs omitted control surfaces in order to minimize drag. They tended to demonstrate instability by going nose down and submerging the upper hull even in calm water. This behavior is analogous to submarine or airship instabilities due to "Munk moments" that grow larger due to the longitudinal pressure distribution on each body of revolution lower hull at an angle of attack. The speed of the onset of instability is directly related to longitudinal metacentric height. A relatively low longitudinal GM is a result of the SWATH geometry.

The aft fins give the stabilizing moment required to counter the Munk moment. Forward fins are destabilizing in pitch, but as in the case with submarines, forward fins are useful as well. The aft fin moment due to location and area must be greater than the forward fin moment to assure stability. Forward fins provide additional damping and give more options to the designer for active control including roll control.

The SSP KAIMALINO, designed in 1970, employs a flap-controlled full-span aft foil connecting the two lower hulls. This foil generated far more control force than the smaller all-movable forward fins, but also generated 1/3 of the total drag of the ship at high speed. The full-span foil at 11.3 meters span and 2.6 meters chord was larger than required for either stability or control in order to have sufficient strength to support itself structurally. The large foil did provide a great deal of damping at speed, but in some following sea conditions it led to large motions when the wave force during long encounter periods pushed up on the foil creating a 'surfing' condition.

Natural Periods

SWATH seakeeping advantages stem partly from long natural periods. Natural periods are a function of waterplane area metacentric height and damping from sources such as the lower hulls and control fins. The advantage of the long natural periods attainable by SWATH ship designs is that resonant conditions leading to large motions and degraded operability will not occur for most wave conditions. The designer desires platforming conditions to minimize pitch and heave motion in head seas. If the natural periods in pitch and heave are about 20 percent greater than the encountered modal period of the seaway this supercritical behavior can be accomplished.

If natural periods are sufficiently long, resonance will only occur in following seas where long wave encounter periods are seen by the ship. The extreme motions associated with resonance will not lead to severe velocities or accelerations since the motions are slow due to the encounter periods being long. This was found to be true in the case of SSP KAIMALINO when operating in following seas at 15.5 knots without automatic motion control. Pitch motions of 20 degrees crest-to-trough were recorded but operability was not adversely affected as long as the propeller remained submerged and the bow was not impacting. Simple automatic or even manual control could easily prevent such extremes of motion. Similar thinking applies to beam sea rolling. A very long roll period assures that resonant conditions will not occur in beam seas under almost all conditions. Resonant conditions that might occur in following seas at moderate speeds can be controlled by automatic motion reduction. Thus the trend in SWATH ship design is to take advantage of long natural periods for good inherent seakeeping in most sea conditions while relying on automatic control and other means to decrease motions when resonant conditions are reached in following seas at moderate speeds.

The SWATH ship designer must also be aware of the interrelationships among the various natural periods. Pitch and roll periods should be separated so that uncomfortable "corkscrewing" motions will not occur in quartering seas. Roll and heave periods and pitch and heave periods should also be separated if possible. The interrelationships among the natural periods are particularly important at low speeds where control fins are not effective in modifying motions. The choice of natural periods may be dictated by low speed behavior, particularly if the ship's mission requires a long duration at low speeds.

Active Control

Active control of motions for SWATH ships differs from that for conventional ships. A SWATH ship has to have fixed fins in order to dampen motions and insure dynamic stability at speed. Thus, the means to actively control motions exists on all designs. The tradeoff in adopting active control, either manual or automatic, is in the need for hydraulic systems to operate the control surfaces. The cost of an autopilot is comparatively insignificant. Movable fins are almost essential on a SWATH since sinkage and trim vary greatly with speed due to the low waterplane area. Movable fins assure level calm water operation at all speeds. Moreover, the smaller waterplane of SWATH ships also allows large changes in the motions with smaller control forces than those required by conventional ships. Control such as adjusting ballast in heavy seas to reduce slams and broaching is also easy and effective for a SWATH design. The active control of motions can be augmented by increasing the size of the fins. Up to a point, increasing fin size also increases the passive damping of the design, thereby reducing motions. These options are open to the designer with the understanding that increased fin size also increases drag.

Control modes can be created for the particular mission of a SWATH design. Platforming control strives to minimize all motions regardless of sea excitation. This is suitable for low sea states and provides an extremely steady platform. The SSP platforming controller demonstrated that such a system can be effectively employed in a SWATH ship of 220 tons. Comprehensive full scale trials results are given by Fein, Ochi, and McCreight (7). The design of the controller is documented by Higdon (8). Figures 6 and 7 show the advantages of the platforming mode of control in virtually eliminating motions of the ship in all sea headings.

The platforming mode can become disadvantageous in the higher and steeper waves associated with extreme seas and building storm seas. In such cases cross-structure impacting and propeller broaching can occur. The platforming mode must be augmented by allowing some contouring of the wave profile in heave to avoid the waves and to minimize the relative motion between the ship and the wave. The successful operation of this control mode was also shown during SSP KAIMALINO trials. Although analog controllers have been used successfully on past SWATH ships, a digital controller offers advantages in flexibility of control strategy and in minimizing control power needed. A consistent digital controller design technique is given by Ware (9).

Low speed reduces the effectiveness of the control surfaces and thus the attenuation of motions. Control surfaces saturate below about 7 knots on the SSP KAIMALINO. Low speed motion characteristics depend on the inherent properties of the geometry rather than control system variables. It is a good practice for the designer to build in natural periods that avoid resonance in most seaways at low speed while relying on the automatic control system to cancel or reduce wave induced motion at high speed.

CONCLUDING REMARKS

This paper briefly gives insight into two critical areas of SWATH dynamics. Maneuvering is important because a proven rudder design approach does not exist and those approaches that have been tried all have implications for reduced propulsive efficiency or increased structural weight. Vertical motion control is of interest because a ship with low vertical motions was the original idea behind SWATH designs; yet good seakeeping performance is not guaranteed, it must be designed into the ship through geometry with active control often necessary.

The results given in Figure 2 demonstrate that a SWATH ship can turn adequately over a range of Froude numbers. The evidence is that a rudder mounted behind the propeller in an extended strut can produce good maneuverability. However the question of rudder design is not closed because a system which did not require the added weight and expense of an extended strut would be preferable. SWATH ships evaluated up to now have exhibited inherently good coursekeeping performance.

The SWATH ship can demonstrate very low motions over a range of wave heights and headings. This is due in part to the low waterplane design which lowers the force generated by the wave and reduces the force required from the fins to control the ship. If the design has adequately long natural periods that postpone resonant conditions to long encounter periods, severe motions will only occur in very high sea states or in following

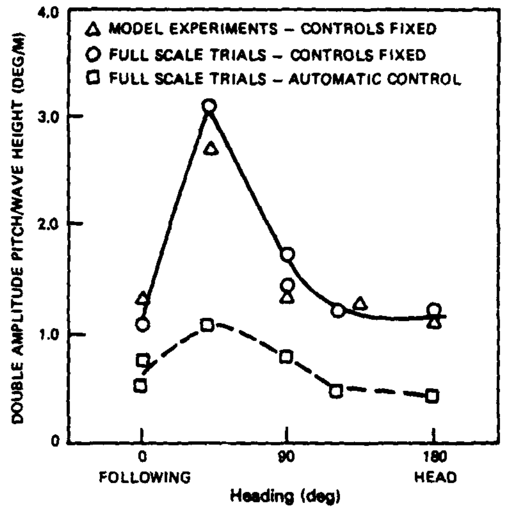


Figure 6. SSP KAIMALINO Pitch Motion at 15.5 knots in a Sea State 5.

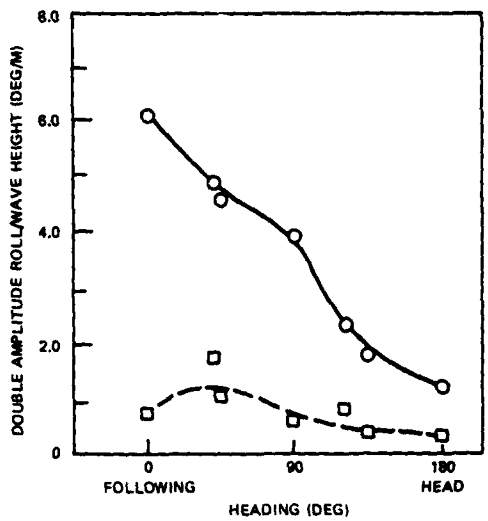


Figure 7. SSP KAIMALINO Roll Motion at 15.5 knots in a Sea State 5.

seas. An automatic controller that allows the ship to contour the waves in such rare conditions while maintaining platforming behavior in other wave conditions will provide all around seakeeping performance far superior to a similar sized monohull.

More research is needed to fully describe the forces acting both in maneuvering and motion control. Another need is for more full scale data on SWATH ships so that a design data base can be built. The SWATH has great potential for a number of missions and it may be only a matter of time before this potential is realized.

REFERENCES

(1) R. Leopold, "A New Hull Form for High-Speed Volume-Limited Displacement-Type Ships," SNAME Spring Meeting, New York, 1969.

(2) T.G. Lang, "Small Waterplane Area Twin Hull (SWATH) Ship Concept and Its Potential," AIAA/SNAME Advanced Marine Vehicles Conference, San Diego, California, April 1978.

(3) R.T. Waters and J.A. Fein, "Maneuverability of SWATH Ships," 19th American Towing Tank Conference, Ann Arbor, Michigan, 1980.

(4) C.J. Hart, R.T. Waters, and J. Hickok, "Rotating Arm Experiment for an Extended Strut SWATH Ship Represented by SWATH-6E," David Taylor Naval Ship R&D Center Report SPD-0698-03, September 1983.

(5) H. Narita, et al, "Design and Full Scale Test Results of Semi-Submerged Catamaran (SSC) Vessels," International Marine Systems Design Conference (IMSDC 182) London, April 1982.

(6) W.C.E. Nethercote, M.D. Miles, and W.D. Molyneux, "Manuevering of SWATH Ships," 20th American Towing Tank Conference, Hoboken, N.J., 1983.

(7) J.A. Fein, M.D. Ochi, and K.K. McCreight, "The Seakeeping Characteristics of a SWATH Ship," 13th ONR Symposium on Naval Hydrodynamics, Tokyo, Japan, October 1980.

(8) D.T. Higdon, "Active Motion Reduction on SSP KAIMALINO in a Seaway," AIAA/SNAME Advanced Marine Vehicles Conference, San Diego, California, April 1978.

(9) J.R. Ware, et al, "SWATH Control Design Using Optimal Techniques," The 6th Ship Control Systems Symposium, Ottawa, Ontario, Canada, 1981.

SWATH VERTICAL MOTION CONTROL USING A FREQUENCY DOMAIN MULTIVARIABLE APPROACH

by F. Caldeira-Saraiva*, D. Clarke* and K. Nicholson+
* British Ship Research Association
+ Admiralty Research Establishment (Haslar)

ABSTRACT

The vertical plane motions of a SWATH (Small Waterplane-Area Twin Hull) craft may be reduced by the addition of horizontal fins to the underwater hulls. Further motion reductions are possible by active control of the fins. Previous attempts to devise a fin control strategy have used a state-space approach and optimal control techniques.

The hydrodynamic behaviour is extremely complicated and is described by differential equations with frequency dependent coefficients. The previously employed state-space approach suffers from a shortcoming that it either over-simplifies the system or demands the introduction of high-order Kalman filters in the feedback loop, in order to observe the necessary system states.

The present paper avoids these difficulties by using advanced frequency-domain multivariable techniques, which take full account of the frequency dependence of the system, without requiring high-order compensation.

Indications of the system performance in irregular seas, stability, and robustness are obtained from the frequency-domain description using Singular Value Decomposition. The controller is synthesised by numerical techniques, to give the closest approach to a Reversed Frame Compensator.

System performance is compared with previously published values of RMS heave and pitch, in typical irregular sea conditions.

INTRODUCTION

The Small Waterplane-Area Twin Hull (SWATH) ship has a large rectangular deck, supported by narrow, surface piercing struts, and submerged lower hulls, as illustrated in Fig.1. This unique hull arrangement which gives a small waterplane-area, enables the maintenance of relatively high speeds in a seaway, because of low motion responses at the more common frequencies of wave encounter.

This combination of a large deck area for a given displacement with low motion responses, makes a SWATH ship attractive as a possible platform to support aircraft operations and other military tasks.

SWATH designs do, however, need horizontal appendages to counteract the Munk Moment on the lower hulls and provide adequate stability in pitch over the desired range of operating speeds in calm water. These necessary appendages have an added advantage that they also act to increase damping in the heave and pitch modes. Moreover, by having the horizontal appendages controllable, further vertical motion reduction is possible. The automatic control of these surfaces has been demonstrated on small ships and improvements in seakeeping performance have been achieved. The power requirements for adequate fin control appeared to be relatively high, so that some optimisation of the control action would be necessary when considering larger SWATH ships.

This paper describes an approach to the problem of optimising the controller design for use on large SWATH ships using multivariable techniques and compares the result with earlier, more simplified approaches.

BACKGROUND AND SCOPE OF THE PAPER

Because of their reduced waterplane areas, SWATH ships are particularly suited for motion control by use of fins. In the last ten years, a large number of studies have dealt with the determination of the size and position of the fins and with the design of the fin controllers (Refs.(1) to (8)). These studies have used, as a model for the motions of the uncontrolled SWATH, simple ordinary second-order differential equations with constant added mass, damping and restoring coefficients, and synthesized, either by frequency domain single-input single-output (Refs.(1), (4) and (5)) or state-space optimal control methods (Refs.(6) to (8)), compensators of the Proportional-Derivative or Proportional-Integral-Derivative types.

Some of these controllers were effectively built into models or full-scale ships and tested (Refs.(9) to (13)). The results of the tests showed improvements that varied amongst the tests and the controllers but were in the best cases for the vertical motions in the region of one-half to one-third in heave and one-third to one-fifth in pitch.

Whilst these studies were proceeding, the knowledge of the hydrodynamic behaviour of SWATHS was also growing. Livingstone (Refs.(14) and (15)) has shown that a good representation of the complex time-domain hydrodynamic behaviour of SWATHS which takes into account frequency dependence can be achieved with high-order systems, typically of order 12.

The idea behind the present paper is to solve the control problem of the determination of the best compensator for the fins, in the framework of a fully frequency dependent model and not in the simplified second-order one. We shall show that taking into account the complexities of the model will allow us to design compensators with better performance and stronger stability and robustness properties.

To achieve this, a frequency-domain multivariable approach is preferable to a state-space one. With a state-space approach we would have to use a high-order description and, since not all system variables would be directly observable, we would need high-order Kalman filters in the feedback loop, thus ending up with extremely complex compensators.

Working directly in the frequency domain on the other hand, is a natural approach, since the description of the hydrodynamic system tends to be obtained in terms of its response to regular trains of waves, i.e. to single-frequency sinusoidal inputs. Thus we have direct knowledge of the open loop transfer function of the system, which is all the basis needed for the multivariable approach. We can therefore save the need to work out the high-order state-space representation of the system.

Having obtained the open-loop transfer function, the controller will be synthesized by numerical methods which allow us to manipulate the open-loop characteristic gains to enhance performance and robustness, while ensuring stability (Refs.(16) to (22)).

THE MULTIVARIABLE APPROACH

The equations of vertical plane motion for SWATH can be written in the form:

$$(A(s)s^2 + B(s)s + C(s))x(s) = T(s)u(s) + q(s)$$

where s is the Laplace variable ($s=j\omega$), A,B and C are the hydrodynamic coefficients, x is the Laplace transform of the output vector whose components are heave and pitch, u is the Laplace transform of the input vector whose components are forward and aft fin deflections, T is a 2x2 matrix (possibly non-constant) and q is the Laplace transform of the forcing vector whose components are heave force and pitch moment.

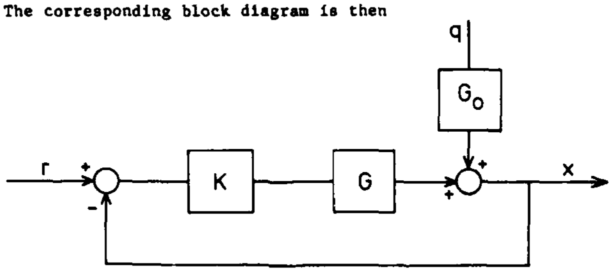
Defining $G_o = (As^2 + Bs + C)^{-1}$ and $G = G_oT$, we can rewrite the equation as:

$$x = Gu + G_oq \qquad (1)$$

The control variable u will be a linear function of output x and a reference vector r. This reference vector can be chosen to be zero if we want to minimize the absolute motion of the ship. We say then that we are aiming at platforming control, and this is the normal mode of operation of the ship. Alternatively, in rough seas, to minimize slamming, the reference input can be taken to be the position of the waves. We say then we are aiming at contouring control.

$$u = -K(x-r) \qquad (2)$$

The corresponding block diagram is then



Replacing the value of u in (1) we obtain, after simplification:

$$x = (I + GK)^{-1} GKr + (I + GK)^{-1} G_oq \qquad (3)$$

Substituting this value in (2) and simplifying we have:

$$u = (I + KG)^{-1} Kr - (I + KG)^{-1} KG_oq \qquad (4)$$

In this paper we shall deal with the platforming problem, for which $r \equiv 0$. We want to have RMS(x(t)) as small as possible compatible with having RMS(u(t)) not too large. If $S(\omega)$ represents the spectrum of the sea then:

$$RMS(x(t)) = \sqrt{\int_0^\infty |x(j\omega)|^2 \, S(\omega) \, d\omega}$$

$$= \sqrt{\int_0^\infty |(I + GK)^{-1} \, G_o q(j\omega)|^2 \, S(\omega) \, d\omega} \qquad (5)$$

and similarly

$$RMS(u(t)) = \sqrt{\int_0^\infty |(I+KG)^{-1} \, KG_o q(j\omega)|^2 \, S(\omega) \, d\omega} \qquad (6)$$

So we want the expression within the integral in (5) to be small for the relevant values of ω whilst at the same time keeping the expression within the integral in (6) suitably small.

REVERSED FRAME CONTROLLER

Suppose we had a SVD[1] of the transfer function G, of the form

$$G = Y \textstyle\sum U^*$$

Suppose now that we could choose a controller K of the form

$$K = U\Gamma Y^*$$

where Γ is a diagonal real matrix. Then:

$$GK = Y \textstyle\sum U^* U \Gamma Y^* = Y \textstyle\sum \Gamma Y^*$$

and

$$KG = U \, \Gamma \textstyle\sum U^*$$

i.e. both GK (the open loop compensated transfer function) and KG have an orthonormal set of eigenframes. The orthonormality of the eigenframes of the open loop compensated transfer function has a large bearing on the robustness of the system (see (22)).

1 Singular Value Decomposition - See Appendix 1.

Apart from ensuring the robustness of the controlled system, this simplifies the calculations of the singular values of $(I+GK)^{-1}$ and $(I+KG)^{-1}K$.

In fact:

$$(I+GK)^{-1} = (I+Y\Sigma\Gamma Y^*)^{-1} = Y(I+\Sigma\Gamma)^{-1}Y^*$$

and

$$(I+KG)^{-1}K = U\ (I+\Gamma\Sigma)^{-1}\ U^*U\Gamma Y^*$$

$$= U\ (I+\Gamma\Sigma)^{-1}\Gamma\ Y^*$$

The singular values of $(I+GK)^{-1}$ are therefore $\frac{1}{1+\sigma_i\gamma_i}$ and those of $(I+KG)^{-1}K$ are $\frac{\gamma_i}{1+\gamma_i\sigma_i}$ where γ_i and σ_i are the diagonal values of Γ and Σ respectively. These values give us all the necessary information on RMS(x) and RMS(u) and thus on the performance of the system.

In terms of the singular values f_i of the open-loop compensated transfer function ($f_i = \sigma_i\gamma_i$), the singular values of $(I+GK)^{-1}$ are $\frac{1}{1+f_i}$ and those of $(I+KG)^{-1}K$ are $\frac{f_i}{\sigma_i(1+f_i)}$.

From what we saw above (equations (5) and (6)), we want $\frac{1}{1+f_i}$ to be as small as possible, in order to minimize motion compatible with keeping $\frac{f_i}{\sigma_i(1+f_i)}$ small, to avoid large fin deflections.

To keep $\frac{1}{1+f_i}$ small we would obviously choose a compensated system with large singular values f_i.

If f_i is large then $\frac{f_i}{\sigma_i(1+f_i)}$ becomes approximately $\frac{1}{\sigma_i}$. For the SWATH, as for most physical systems, σ_i is comparatively large for small values of ω, and so the fin action can be kept small. As $\omega\to\infty$, however, σ_i will tend to zero so fin action can only be kept small for high frequencies if $f_i\to 0$ as $\omega\to\infty$. This will contradict the earlier requirement that $\frac{1}{1+f_i}$ be as small as possible, but for large values of ω, $S(\omega)$ will be small, so there will be little effect on the RMS.

The policy for selecting the optimal compensated system is then clear: it should have singular values as high as possible near $\omega=0$ and over the significant wave band. This ensures that heave and pitch are small, whilst at the same time (since σ_i is large) does not induce large fin motions. Then, as ω increases, the singular values should decrease sharply to guarantee small fin actions. In other words, the ideal system should incorporate a number of low-pass filters to provide the necessary high magnitude at low frequencies together with low magnitude at high frequencies. To preserve stability, some lead action might be necessary, so a number of zeros could also be incorporated.

After having selected a suitable open-loop compensated transfer function, one can synthesize by numerical methods (as detailed in Appendix 2 - see also Ref.(22)) the controller K that brings the system to the required form and best approaches a reversed frame compensator.

THE CONTROL OF SWATHS

We shall now give an example of the application of the method described above to a SWATH of the 6A type. (See Ref(6), for example, for the characteristics of a SWATH 6A.)

In Figs.2 to 6 we see the indices of stability and performance of the uncompensated SWATH 6A. Figure 2 shows the Nyquist diagram for the two eigenvalues of the system. Figures 3 and 4 show the corresponding Bode diagrams for magnitude and phase. Figure 5 shows the misalignment which is a measure of the deviation of the eigenframes of the system from orthonormality. The maximum theoretical value of the misalignment for a two-input two-output system is $\sqrt{2} \approx 1.4$.

Figures 6 to 8 show the Quasi-Nyquist diagrams. From Figures 6 and 7 we can see the magnitudes of the singular values which are responsible for performance The RMS values for heave and pitch (displacement, velocity and acceleration) for the uncompensated SWATH at 20 knots in head sea with a sea state 6, obtained in the frequency domain with the ITTC spectrum were:

	Heave		Pitch
Displacement (ft)	1.87	Displacement (deg)	.46
Velocity (ft/s)	1.41	Velocity (deg/s)	.48
Acceleration (g)	.036	Acceleration (deg/s^2)	.58

Figures 9 to 16 show the stability, performance and robustness indices for an ideal two-by-two system with the transfer function $\frac{10s^3+204s^2+80s+8.6}{10s^4+31s^3+31s^2+20s+7.3}$ I. As can be seen in Figures 13 (Nyquist diagram) or 14 (Bode magnitude diagram) the magnitude of the singular values is very high at medium frequencies but decreases very steeply to low values at high frequencies. The misalignment of the system is nil, Fig.12 only showing the computing inaccuracies.

The RMS values for heave, pitch, forward fin deflection and after fin deflection for the ideal system were:

Heave		Pitch		Fore Fin Deflection	Aft Fin Deflection
Displacment (ft)	.1	Displacement (deg)	.04	6.4	4.2
Velocity (ft/s)	.09	Velocity (deg/s)	.07	10.2	5.8
Acceleration (g)	.004	Acceleration (deg/s ^{2})	.15	32.3	12.2

There is therefore an improvement of 17 times in heave and 12 times in pitch with very acceptable fin actions. From the Rayleigh distribution, twice the RMS represents the one-third highest or significant average while 2.546 times the RMS represents the one-tenth highest average. The values for the fin deflections are:

	Fore fin deflection	Aft fin deflection
One-third highest displacement	12.9	8.4
One-tenth highest displacement	16.4	10.5
One-third highest velocity	20.4	11.5
One-tenth highest velocity	26.0	14.8

which compare very well with the maximum values used in practice for deflection and rate of deflection (28 degrees and 40 degrees/s respectively).

The rational near-reversed-frame precompensator obtained by best fitting was (after simplification):

$$K = \frac{1}{s^2+5s+2}\begin{bmatrix} 43 + 425s + 1260s^2 & -2 + 38s - 263s^2 \\ 9 + 75s + 292s^2 & 1 + 29s + 131s^2 \end{bmatrix}$$

Figures 17 to 24 show the indices of stability, performance and robustness for the compensated system.

Figure 20 shows that the system is very nearly orthonormal (misalignment is below 0.4 for the relevant range of frequencies). This means that the system's singular values are not too sensitive to variations in the system's parameters. This is very important in the present problem because of the approximations made in the calculation of the hydrodynamic coefficients.

Figures 21 to 23 show the corresponding Quasi-Nyquist diagrams which, because of the near orthonormality of the system, are very similar to the eigenvalue diagrams. Finally Fig. 24 shows the robustness of the system i.e. the margins of stability at each frequency. We see that it is always greater than .9 which means that a change in the gains of up to 90% would not affect stability.

The RMS values for this compensated system are:

	Heave		Pitch	Fore fin deflection	Aft fin deflection
Displacement (ft)	.11	Displacement (deg)	.07	6.4	3.9
Velocity (ft/s)	.10	Velocity (deg/s)	.09	9.8	5.2
Acceleration(g)	.004	Acceleration (deg/s ^{2})	.16	31.6	10.6

The degradation from the ideal system is therefore not too large. Improvements of 17 times on heave and 7 times on pitch are still obtained and the third highest and one-tenth highest values for fin actions are:

	Fore fin deflection	Aft fin deflection
One-third highest displacement	12.7	7.7
One-tenth highest displacement	16.2	9.8
One-third highest velocity	19.6	10.4
One-tenth highest velocity	25.0	13.2

i.e. still within acceptable limits.

For comparison purposes we shall now review the state feedback optimal controller presented in Ware and Scott (7), again for platforming at 20 knots in head sea with sea state 6.

Figures 25 to 32 present the indices of stability, performance and robustness for the system with this precompensator. We see in Figs. 29 and 30 that singular values at high frequencies are low as they would have to be to guarantee small fin actions. However, the singular value associated with heave still has a modest value at low frequencies, as compared with the reversed frame compensated one (Figs. 13 and 14), though it is better than that for the uncompensated system (Figs. 6 and 7). The same figures show also that there is no improvement in pitch. Thus, while fin actions will be suitably low, heave and pitch will be higher than the ones we obtained above. In fact the RMS values, obtained in the frequency domain were:

	Heave		Pitch	Fore fin Deflection	Aft fin Deflection
Displacement (ft)	.50	Displacement (deg)	.48	3.2	3.6
Velocity (ft/s)	.42	Velocity (deg/s)	.45	3.1	3.4
Acceleration (g)	.013	Acceleration (deg/s ^{2})	.57	4.2	4.3

which represent an improvement of 3.7 times on heave and none on pitch (compared to 17 times and 7 times before).

The values given by Ware and Scott for the RMS values are:

	Heave		Pitch	Fore fin Deflection	Aft fin Deflection
Displacement (ft)	.633	Displacement (deg)	.545	4.189	4.609
		Velocity (deg/s)		4.681	4.609

which are similar to the above.

A perfect correspondence could not be expected since Ware and Scott's values were obtained by simulation with a simplified model while the values given above were obtained in the frequency domain. Ware and Scott's values for the uncompensated SWATH are given as:

	Heave		Pitch
Displacement (ft)	2.251	Displacement (deg)	.708

which again are similar but not exactly the same as those we obtained above.

Performance is therefore worse for this state feedback controller. Furthermore, if we look at Fig. 28 we see that the compensated system is very far from achieving othonormality (remembering that the maximum theoretical value for misalignment is 1.4). This means that the system's singular values are sensitive to changes in the system's parameters which suggests that performance in practice may be very different from that predicted above. Again, as we see from Fig. 32, the system's stability margin is lower than that of the multivariable controller. A change of 20% in the system's parameters may make it unstable.

CONCLUSION

We conclude therefore that the compensated system obtained by frequency domain multivariable techniques compares very favourably with that obtained by the state-feedback approach. Moreover, whilst the other approach does not concern itself with orthonormality and robustness, the Quasi-Classical approach ensures automatically that the compensated system is nearly orthonormal and thus has low sensitivity to perturbations, and has good stability margins.

Finally, while the state variable approach demands from the designer the specification of weighting matrices for state and control deviations, which is not easy to do or to have a feeling for, the Quasi-Classical approach works only with the usual frequency domain concepts and provides results directly in terms of RMS values for which the designer has an immediate feeling.

ACKNOWLEDGMENT

The authors express their thanks to the Ministry of Defence and the Research Council of the British Ship Research Association for permission to publish this paper.

REFERENCES

(1) J.A. Fein and J.P. Feldman, "Controllability of the Stable Semi-Submerged Platform", 3rd Ship Control Systems Symposium, Paper IXB-3, 1972.

(2) C.M. Lee and M. Martin, "Determination of Stabilizing Fins for SWATH Ships", 4th Ship Control Systems Symposium, Vol.4, 1975, p.58.

(3) C.M. Lee and R.C. Curphey, "Prediction of Motion, Stability and Wave Load of Small-Waterplane-Area, Twin-Hull Ships", SNAME Transactions, Vol.85, 1977, p.94.

(4) C.M. Lee and K.K. McCreight, "Investigation of Effects of Activated Fins on Vertical Motion of a SWATH Ship in Waves", DTNSRDC/SPD-063-01, 1977.

(5) D.T. Higdon, "Active Motion Reduction for a 3000-Ton SWATH Ship Underway in Regular Head and Following Waves", Seaco Inc., Report No.80-05-01, 1980.

(6) J.R. Ware, J.F. Best and V.A. Scott, "Applications of Optimal Control Techniques to SWATH Vertical Plane Control", ORI Tech. Report TR-1682, 1980.

(7) J.R. Ware and V.A. Scott, "Comparison of Platforming and Contouring Automatic Control Systems for SWATH Craft - An Optimal Control Approach", ORI Tech. Report TR-1790, 1980.

(8) J.R. Ware, J.F. Best, V.A. Scott and K.K. McCreight, "SWATH Control Design using Optimal Techniques", Proc. 6th Ship Control Systems Symposium, Vol.3, 1981.

(9) J.A. Fein, K.K. McCreight and J.A. Kalio, "Seakeeping of the SSP KAIMALINO", AIAA/SNAME Advanced Marine Vehicles Conference, Paper 78-739, 1978.

(10) D.T. Higdon, "Active Motion Reduction on SSP KAIMALINO in a Seaway", AIAA/SNAME Advanced Marine Vehicles Conference, 1978.

(11) J.A. Fein, M.D. Ochi and K.K. McCreight, "The Seakeeping Characteristics of a Small Waterplane Area, Twin-Hull (SWATH) Ship", 13th Symposium on Naval Hydrodynamics, Vol.3, 1980.

(12) J.A. Fein, "Seakeeping and Motion Control Trials of SSP KAIMALINO in Sea States 4 and 5", DTNSRDC - 81/015, 1981.

(13) H. Narita, T. Mabuchi, Y, Kumitake, H. Nakamura and M. Matsushima, "Design and Full Scale Test Results of Semi-Submerged Catamaran (SSC) Vessels", First International Marine Systems Design Conference, Paper No.11, 1982.

(14) W. Livingstone, "Generalized Nonlinear Time Domain Motion Predictor for SWATH", DTNSRDC/SPD-0857-01, 1979.

(15) W. Livingstone, "A Nonlinear Frequency Domain Simulation for SWATH Craft", DTNSRDC/SPD-0893-01, 1980.

(16) A.G.J. MacFarlane and B. Kouvaritakis, "A Design Technique for Linear Multivariable Feedback Systems", International Journal of Control, Vol.25, 1977, p.837.

(17) J.M. Edmunds and B. Kouvaritakis, "Extensions of the Frame Alignment Technique and Their Use in The Characteristic Locus Design Method", International Journal of Control, Vol.29, 1979, p.787.

(18) J.M. Edmunds, "Control System Design and Analysis using Closed-loop Nyquist and Bode Analysis", International Journal of Control, Vol.30, 1979, p.773.

(19) J.C. Doyle and G. Stein, "Multivariable Feedback Design: Concepts for a Classical/Modern Synthesis", IEEE Transactions on Automatic Control, Vol.AC-26, 1981, p.4.

(20) C.A. Desoer and M.J. Chen, "Design of Multivariable Feedback Systems with Stable Plant", IEEE Transactions on Automatic Control, Vol.AC-26, 1981, p.408.

(21) I. Postlethwaite, "A Design Procedure for Linear Multivariable Feedback Systems", Report CVED/F-CAMS/TR-210, Engineering Department, Cambridge University, 1981.

(22) Y.S. Hung and A.G.J. MacFarlane, "Multivariable Feedback: A Quasi-Classical Approach", Springer-Verlag, 1983.

APPENDIX 1

SINGULAR VALUE DECOMPOSITION

Given any complex matrix A, if we define its conjugate by A^*, then for any vector x:

$$|Ax|^2 = x^* A^* Ax$$

and we define the gain of the matrix for such vector x by

$$\frac{|Ax|}{|x|} = \sqrt{\frac{x^* A^* Ax}{x^* x}}$$

For each different x a different gain will result. If we want to find the maximum and minimum gains for a vector of given modulus, we form the Lagrangean $L = |Ax|^2 - \lambda|x|^2$ and differentiate it to obtain the condition $A^*Ax = \lambda x$, which means that the maximum and minimum gains are obtained when x is aligned with two of the eigenvectors of A^*A. Since A^*A is necessarily Hermitian and positive-definite, its eigenvalues will be real and positive. If we denote by σ^2 one of them and if x is aligned with the respective eigenvector then:

$$\frac{|Ax|}{|x|} = \sqrt{\frac{x^* \sigma^2 x}{x^* x}} = \sigma$$

We call the values σ, the singular values of A. Among the singular values thus there will be the maximum and minimum gains of the matrix.

By definition of eigenvalues and eigenvectors we can write:

$$(A^*A)U = U\Sigma^2 \qquad (1)$$

where Σ^2 is the diagonal matrix of the eigenvalues of A^*A and U is a matrix whose columns are the eigenvectors of A^*A. Moreover we can always choose U so that it is orthonormal, i.e.

$$U^*U = UU^* = I$$

Multiplying (1) by A we have

$$(AA^*)AU = AU\Sigma^2$$

or also

$$(AA^*)\,(AU\Sigma^{-1}) = (AU\Sigma^{-1})\,\Sigma^2$$

If we define $Y = AU\Sigma^{-1}$, then Y is the matrix whose columns are the eigenvectors of AA^* and it is simple to check that Y is orthonormal.

Finally we also have

$$Y\Sigma U^* = (AU\Sigma^{-1})\ \Sigma U^* = A$$

and so every non-singular complex matrix A has a decomposition of the form

$$A = Y\ \Sigma\ U^*$$

where U is the matrix of the eigenvectors of A^*A, Y the matrix of the eigenvectors of AA^*, and Σ a real diagonal matrix of the singular values. This decomposition is called a singular value decomposition, and U and Y are respectively the right and left eigenframes.

APPENDIX 2

THE SELECTION OF A NEAR-REVERSED FRAME COMPENSATOR

If the uncompensated transfer function G has a singular value decomposition

$$G = Y \Sigma U^*$$

then an ideal reversed-frame precompensator will be of the form

$$K = U T Y^*$$

producing an ideally compensated system:

$$Q = GK = Y \Sigma T Y^*$$

Working in reverse order, if Θ represents the desired orthornormal ideal system, then

$$Q = Y \Theta Y^*$$

will be the system's compensated ideal, which has the same singular values as Θ and is orthonormal.

We shall fit a compensator K of the form:

$$K(s) = \frac{1}{d(s)} N(s)$$

where d(s) is a given common denominator and N(s) a matrix of polynominals. If we put:

$$S(s) = \begin{bmatrix} 1 & s & s^2 \dots & s^d & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 & s & s^2 \dots & s^d \end{bmatrix}$$

and

$$N = \begin{bmatrix} a^o_{11} & a^1_{11} & \cdots & a^d_{11} & a^o_{21} & a^1_{21} & \cdots & a^d_{21} \\ a^o_{12} & a^1_{12} & \cdots & a^d_{12} & a^o_{22} & a^1_{22} & \cdots & a^d_{22} \end{bmatrix}^T$$

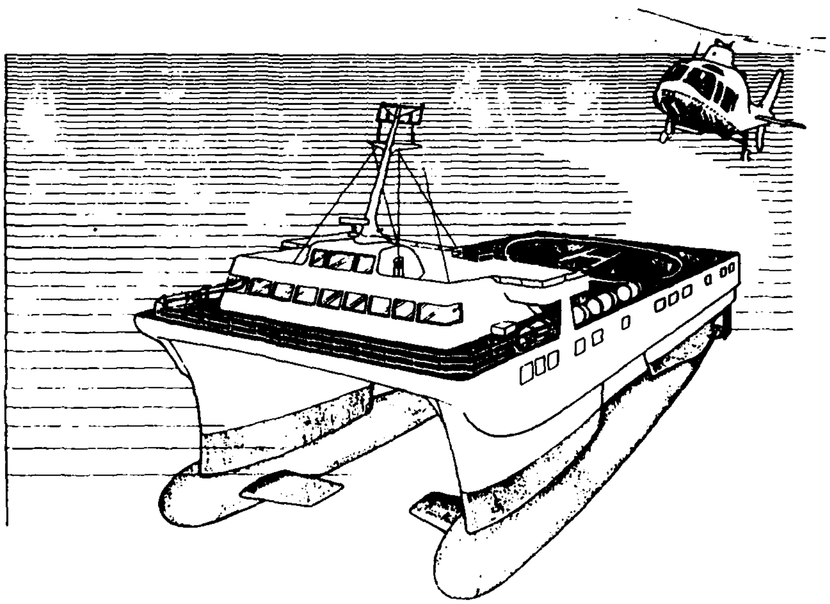


Figure 1. Artist's Impression of a SWATH Vessel

Uncompensated System

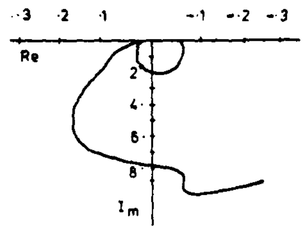


Figure 2 Characteristic Gains Nyquist Diagram

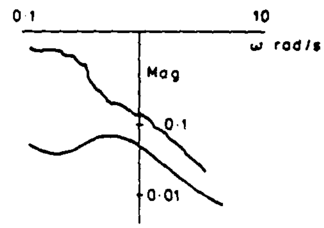


Figure 3 Characteristic Gains Bode Magnitude Diagram

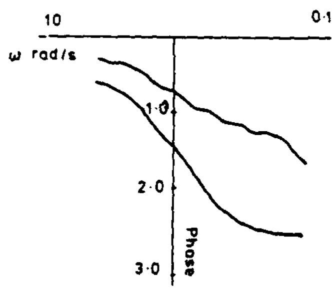


Figure 4 Characteristic Gains Bode Phase Diagram

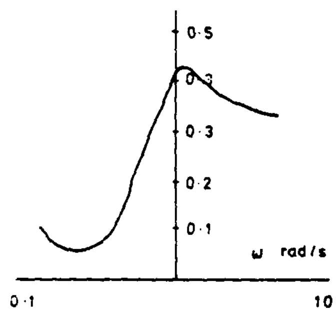
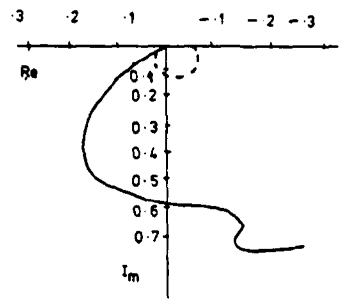


Figure 5 Misalignment

Uncompensated System (Cont.)



gure 6 Quasi - Nyquist Loci
Nyquist Diagram

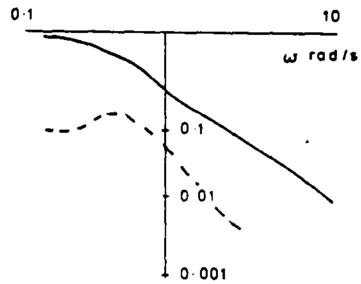


Figure 7 Quasi - Nyquist Loci
Bode Magnitude Diagram

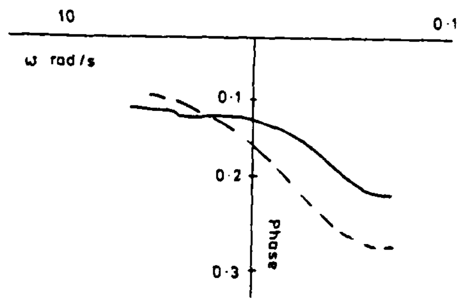


Figure 8 Quasi - Nyquist Loci
Bode Phase Diagram

Ideal System

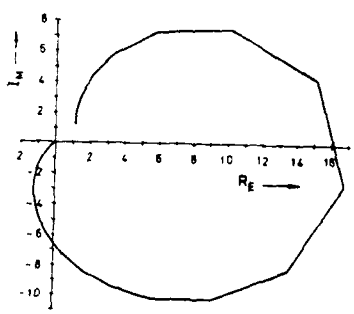


Figure 9 Characteristic Gains Nyquist Diagram

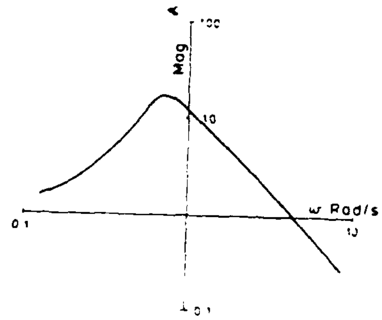


Figure 10 Characteristic Gains Bode Magnitude Diagram

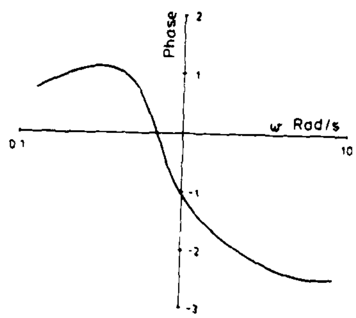


Figure 11 Characteristic Gains Bode Phase Diagram

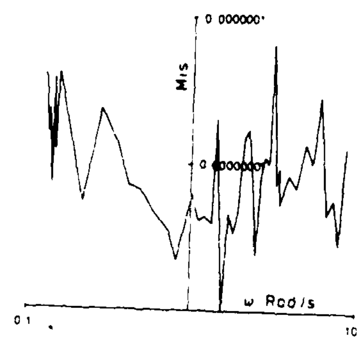


Figure 12 Misalignment

Ideal System (Cont)

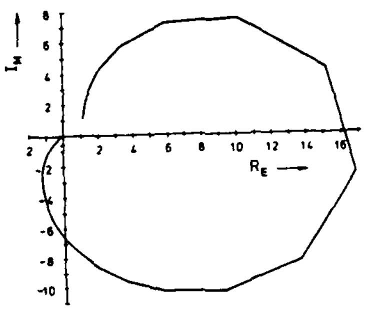


Figure 13. Quasi-Nyquist Loci
Nyquist Diagram

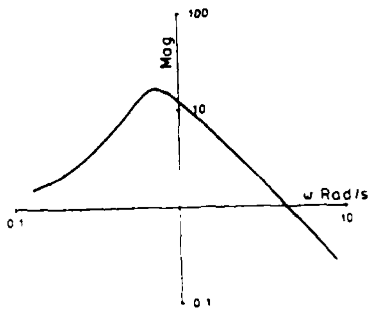


Figure 14 Quasi-Nyquist Loci
Bode Magnitude Diagram

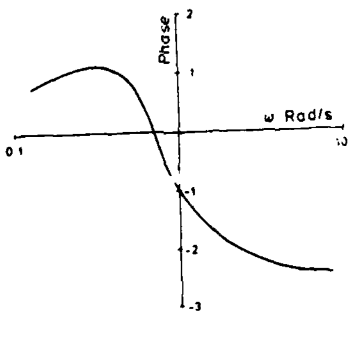


Figure 15 Quasi-Nyquist Loci
Bode Phase Diagram

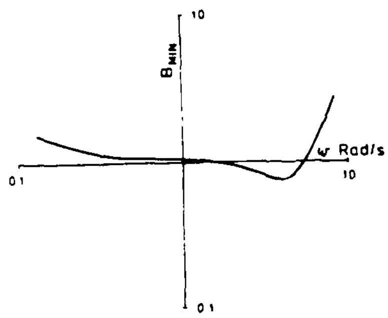


Figure 16 Robustness

Compensated System

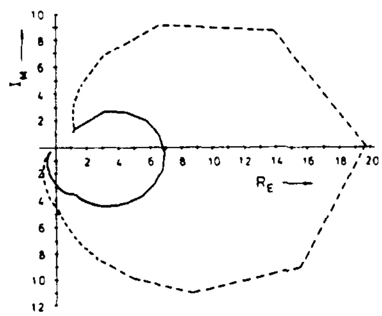


Figure 17 Characteristic Gains
Nyquist Diagram

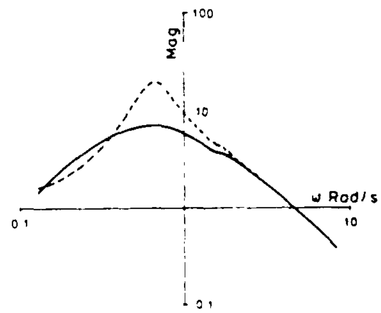


Figure 18 Characteristic Gains
Bode Magnitude Diagram

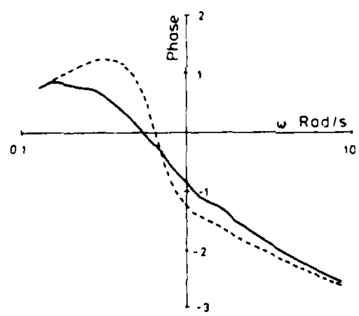


Figure 19. Characteristic Gains
Bode Phase Diagram

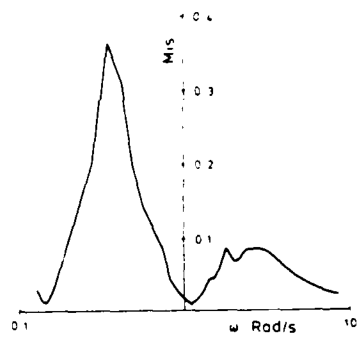


Figure 20 Misalignment

Compensated System (Cont.)

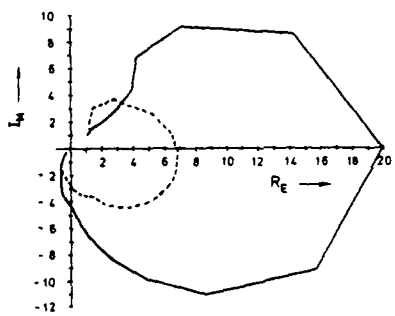


Figure 21. Quasi-Nyquist Loci
Nyquist Loci

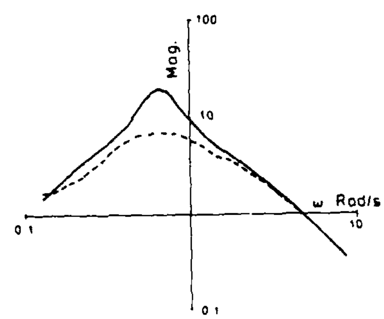


Figure 22 Quasi-Nyquist Loci
Bode Magnitude Diagram

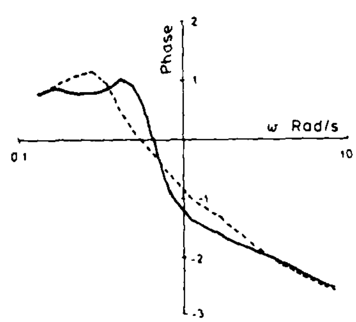


Figure 23 Quasi-Nyquist Loci
Bode Phase Diagram

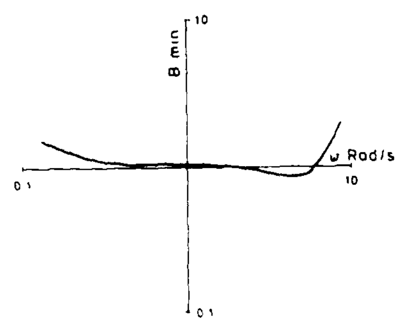


Figure 24 Robustness

Optimal Control System

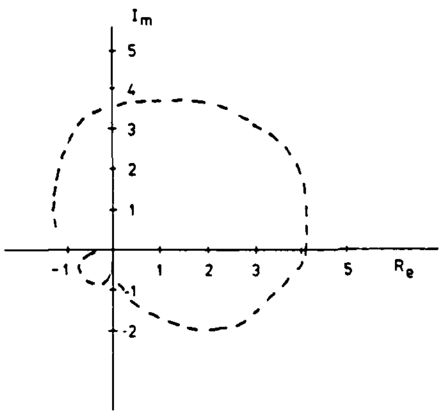


Figure 25 Characteristic Gains Nyquist Diagram

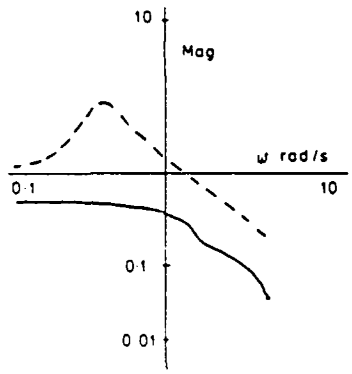


Figure 25 Characteristic Gains Bode Magnitude Diag

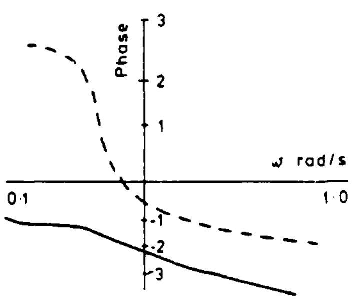


Figure 27 Characteristic Gains Bode Phase Diagram

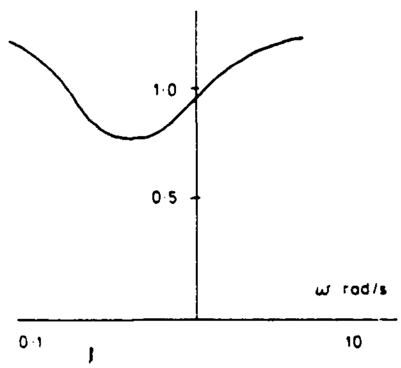


Figure 25 Misalignment

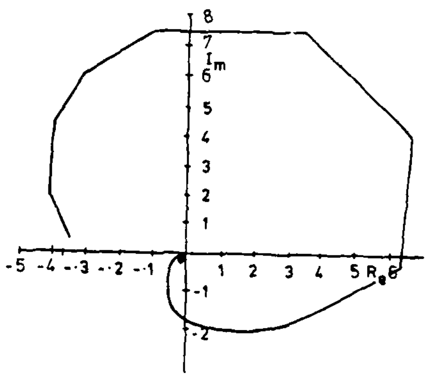


Figure 29 Quasi - Nyquist Loci
Nyquist Diagram

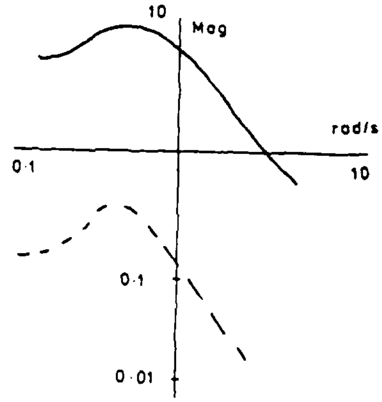


Figure 30 Quasi - Nyquist Loci
Bode Magnitude Diagram

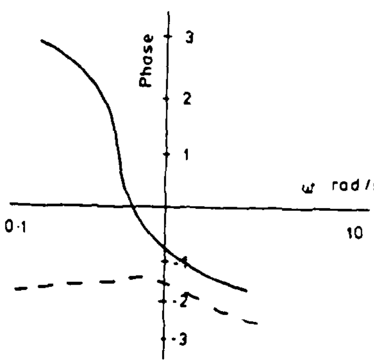


Figure 31 Quasi - Nyquist Loci
Bode Phase Diagram

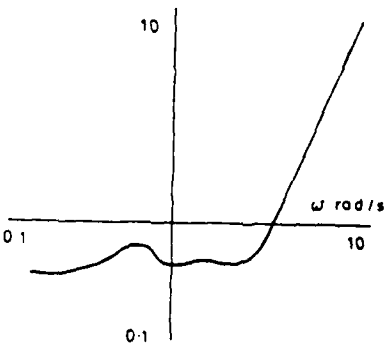


Figure 32 Robustness

OPTIMAL DESIGN OF DIGITAL PROPULSION CONTROL
FOR THE SPANISH CARRIER SC-175: ASSOCIATED TRENDS

by C. Cassel and E. Pollak
Simulation and Control Systems Department
General Electric Company
Daytona Beach, Florida
and
Baudilio Sanmartin
Bazan Naval Shipyard
El Ferrol, Spain

ABSTRACT

The new Spanish aircraft carrier, the Principe De Asturias, (R-11), (figure 1), has a displacement of 15,000 tons (fully loaded) and is powered by two LM-2500 gas turbine engines, with a single controllable-reversible pitch propeller (CRP). Its conservative power-to-weight ratio, along with its military performance objectives, presented a special challenge to the design of its propulsion control system. The scheduled-adaptive control technique developed for the Oliver Hazard Perry class frigate, and now well proven by fleet experience, was extended to provide the required degree of optimization in some performance regions, such as the crash stop.

The identification of the controller as "scheduled-adaptive" is qualified, and its inherent design problems are outlined. In keeping with the Seventh Symposium's special interest in the practical problems of realization of a successful digital control system, the relevant background and experience of the Perry and SC-175 controller design activities are reviewed, and some details of the design methods and outcomes are presented. Performance predictions for several critical maneuvers, including the crash stop, are presented. Some generalizations on the use of the design method, and speculations on its further potentials are offered.

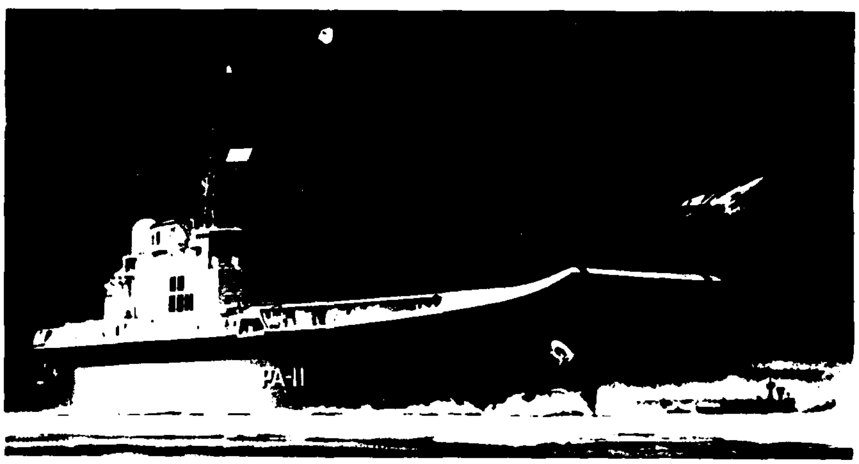


Figure 1. Principe De Asturias (R-11)

INTRODUCTION

It is now ten years since simulation studies were started for the development of a new type of COGAG-CRP propulsion control for the U.S. Navy Perry class frigate. The new approach put great emphasis on the exploitation of the capabilities of the digital computer as an in-line controller. The lead ship, the Oliver Hazard Perry, was launched for sea trials from Bath, Maine, USA. The sea trials were completed in record time, and Admiral John D. Bulkeley announced that the Perry was "considered fully satisfactory in all respects for acceptance." At this writing the Perry class fleet has accumulated over 100 ship-years of highly successful propulsion control operations, with a nearly perfect record for reliability of the digital controller hardware and firmware. A major advance in reliability was achieved in conjunction with a considerable increase in performance flexibility (1).

The Spanish aircraft carrier, the SC-175, or the Principe De Asturias (R-11), was launched at El Ferrol Del Caudillo in northwestern Spain on 22 May 1982. The building of the SC-175 was a part of a package plan including the building of three Perry frigates, also at the Bazan Shipyard at El Ferrol (2). The Gibbs & Cox Company, ship architects, was selected by Bazan to support the design of the SC-175. The General Electric Company at Daytona Beach, Florida, USA, was selected, as it had been for the Perry frigate, to design ship controls systems including the COGAG-CRP digital controller. With the requirement for maximum commonality of equipment as practicable, and with the same design team connections, including General Electric Company personnel at the Aircraft Engine Group (AEG) Evendale, Ohio, USA engine plant, the new challenges of the Dynamic Response Analysis (DRA) and controller design were approached as a continuation of a series of development studies which had been undertaken for the design of the Perry controller. These unique circumstances, and the evident success of the Perry controller approach, together with the special interest and supervision of the DRA by S. Baudilio Sanmartin, the chief Bazan engineer assigned to the control systems procurement, favored extended dependence upon the Perry controller design assumptions.

The Perry controller design concepts were synthesized pragmatically from investigations into the aspects of the problem at practical levels, using working relationships and contacts with experts dedicated to various critical engineering areas. The synthesis tool was digital simulation, including the full-scale, detailed and full-parameter model, for verification of simplified and extended simulation models. The Daytona Beach General Electric Company component is the Simulation and Controls Systems Department (GE/SCSD), which has resources that stimulate and accommodate the approach. While there is general consensus on the importance of the use of the full parameter nonlinear simulation model as a design tool, in one way or another, the approach used for the SC-175 represents a departure from others. Section 2 attempts to identify and clarify that approach in relation to global theory and to other concepts and approaches that have been advocated.

Using the perspective that this provides, the bulk of this paper relates the applicable Perry and SC-175 controller design experiences. Actual design methods and design data are presented, to clarify the approach and to deal at the practical level. Some performance predictions are presented, including the crash stop, to show how they may differ from those in other approaches, and to what degree the characteristics of true optimization are achieved.

The main leverage of the approach presented is considered to be the advanced status and continued rapid progress in capabilities of generally available computers and their support software, particularly as a design tool. The ship development cycle is so long in relation to the pace of this advance, as was strikingly experienced in going from the Perry to the SC-175 activity, that general conclusions and speculations on the further use and improvement of the scheduled-adaptive design approach for the COGAG-CRP and other ship propulsion systems must be projected into the expected future resource environment, if they are to be useful. This paper concludes by summarizing what seem to be the significant aspects of the Perry and SC-175 controller design experiences into a few such conclusions and speculations.

THEORETICAL BACKGROUND

The Proceedings of an international symposium on methods and applications in adaptive control, held at Bochum in 1980, contain a number of papers relating aspects of adaptive system theory and design relevant to our interest. A survey paper by Parks et. al. (3) divides adaptive control systems into three structures, using the open loop structure as a third category. "Gain scheduling" is identified as a type of open loop adaptive control. The scheduling system is adaptive but not self-adaptive. In the open loop system the decision process is reduced to a fixed mapping of the process parameters to the controller parameters. The original decision process is already realized in the design phase. The type structure is said to be in widespread use and vogue today, allowing one to tune a wide range of controllers using a manifold of popular on-line process identification methods. The survey paper was limited to the review of aircraft control systems, process control, and electrical drives. Confidence in the design is obtained on the basis of a "good" (adequate) knowledge of the actual process dynamics. Parks points out that the identification techniques must be robust and reliable, as uncertainties in the process parameters are not taken into account.

A paper by Tiano et. al. presented at the Sixth Ship Control Systems Symposium (4) quotes Tsypkin's definition of adaptive control (5):

> "Adaptation is the process of changing the parameters, structure and possibly the controls of a system on the basis of information obtained during the control period, so as to optimize - from one point of view or another - the state of the system, when the operating conditions are either incompletely defined initially or changed."

The Tiano paper compares various adaptive control techniques used for ship steering and course keeping, and treats the practical problems of applications, with insights into successful approaches, especially in relation to the use of simulation for design. He stresses the need for understanding the basics of the dynamics and of the modeling craft, and for the associated need for a human network connecting sources of expertise and data. They must constitute an integrated activity of people who "know the physics of the process, the ins and outs of automatic control,and of people who know how to make a flexible computer program and how to set-up and run an adequate simulation run program." This indicates an approach interested in partitioning performance into regions in which separate controllers or algorithms will be postulated on the basis of knowledge of the system, then tested and refined via simulation. Not just gains, but algorithms are switched, using robust and reliable identifiers, which may include any information available during the control cycle. Moreover, different types and degrees of "optimization" can be applied for different performance regions. The Tiano paper offers vital insights into good simulation practice, noting that small errors in the model will lead to bounded errors in the response of the ship, due to the adaptive design, but that insufficient knowledge of the system or a too simple model may fail to cover significant performance regions.

Tanner et. al. (6) reviewed the use of simulation for warship control system design and development at the Sixth Symposium. The need for full and detailed simulation was recognized, with a special interest in real-time simulation for equipment testing. The paper explained a program for accumulating and extending resources for simulation testing and identified existing and expected application areas. A current trend toward advancement from earlier control concepts involving closed loop control of shaft speed, or of pitch rate, or of shaft torque was noted. Open loop control with design simulation used to select best command rate limits and the use of identifiers such as anticipation of pitch reversal has proven to be successful. Simulation for test, as well as design, and for support of trials and operations is in development and use. The SC-175 design approach is compatible with this program, particularly in recognizing the need for open loop control, but tends to extend it by partitioning performance regions to handle respective performance envelopes.

Kidd et. al. presented a very interesting projection on the subject controller design problem at the Sixth Symposium (7). By using multivariable control theory with linearization, an analysis revealed in relatively precise terms the great variation of response, showing that "the system is extremely nonlinear with variations in both gains and dynamic terms of the order of a hundred to one." Using a theoretical approach, the potentials for multivariable control with optimization techniques, as opposed to the typical performance of the conventional controller were illustrated. An adaptive multivariable compensator would be developed by mapping gains from design simulations into the compensator design. The gains would be scheduled in response to feedback of state data for uncoupled, multiloop, linearized control, accommodating optimization. Again, the use of detailed, nonlinear simulation in the design stage is found to be necessary. In fully practical system designs various other constraints must be considered, in various performance regimes. The powerful tools used for the analysis, however, have considerable promise in supporting the further development of the scheduled-adaptive approach. Again, they are tools that are coming from the development of the digital computer and its support software that can be used to make up for the lack of global theory (nonlinear control).

We have arbitrarily named the Perry and SC-175 approach the "scheduled-adaptive" approach for the purposes of this paper. Its basic concept is to divide performance into regions (partition state space) as inspired by understanding of the dynamics of the system, and to fashion control algorithms for each region, then test and refine them using simulation. In each control cycle the controller is to sense which region should be applied, and the algorithms for that region provide incrementations for command outputs as based on the current system state and the state of the inputs to the controller (ship speed demand, mode to be used, etc.). Identifiers to be used for switching control regions are to be robust and reliable (shaft speeds has gone above a given level, a given pitch has been reached, etc.). Conservative closed loop control of shaft speed is offered as an operator option, but the primary steady (cruising) mode is open loop. The control algorithms use inputs to the controller from calibration settings, and interpret the calibration bias for use at the current operating point. Major performance regions, e.g., the crash stop, are partitioned into stages which further specialize algorithm selections to "shape" responses toward performance objectives for the parameter trajectories, while satisfying any number of defined constraints. Interactive design simulations are used for the shaping, and for systematic validations showing that extreme or worst cases satisfy requirements, and that other cases within particular envelopes are adequate.

The open loop adaptive control system tends to have inherent stability. Some of the algorithm techniques applied use a closed loop control of shaft speed, temporarily, to make them robust over cases and with unknown disturbances, and an automatic speed (shaft feedback) mode is used with a fixed idling shaft speed in the SC-175 design. Studies of performance with wave actions at mild and extreme sea states have not revealed any stability problems. Moreover, the system provides means to tune down gains for the Proportional-Integral (PI) compensator and to use a "heavy seas damping" switch to reduce them to nearly negligible levels. The design also features a monitoring and override control for bounding shaft speed. If shaft speed goes higher or lower than set speeds the "cutback" or "boost" term is accumulated as negative or positive, respectively, and added to the regular power command. When these terms are not zero and shaft speed is within bounds, they are degraded toward zero at significantly lower rates than their accumulation rates. The net effect of the cutback operation in very heavy seas is to prohibit high level power commands (which would not occur within reasonable operations).

Although the design approach described in this paper can be used to achieve relatively optimal control, as was needed for some of the SC-175 performance, (Donneley has described such optimization for the crash action in reference 11), it did not come about primarily in an effort to optimize performance. Rather, it evolved through special efforts to exploit the digital controller in eliminating problems that were being experienced in the COGAG-CRP controller design field, as cited in Tanner (6), such as resulted from closed loop control of shaft speed or of shaft torque. Another such problem which tended to dominate the Perry and SC-175 controller design was the need to assure against overthrusting the CRP blades (11). Other problems were involved, so that the general trend of the design effort was toward methods for coping with any variety of constraints, i.e., the partitioning of performance and specializing of controls within regions. Let us now turn to an account of some of the design experiences.

THE PERRY CONTROLLER DESIGN EXPERIENCE

GE/SCSD became the controls systems contractor for the FFG-7 lead ship and fleet under the Bath Iron Works Shipbuilders, with the Gibbs & Cox Company (G&C) as Design Agent in 1973. Simulation studies that had been done by the U.S. Navy at the Naval Research and Development Center at Annapolis, Maryland, USA, in collaboration with GE/AEG, that furnished an early full-parameter, nonlinear simulation program for the LM-2500 gas turbine engine, constituted a background for early modeling and simulation work. The Annapolis work had addressed a ship design similar to that of the FFG-7, and was reported into the literature by C. J. Rubis (8,9), now of Propulsion Dynamics, USA. GE/SCSD established a direct working relationship with GE/AEG and began to use an up-to-date version of the full parameter engine simulation program. Performance requirements and ship system data were provided by G&C. After a report on the Dynamic Response Analysis (DRA) study completed at GE/SCSD in the fall of 1973, it was determined that a series of phased studies should be used to investigate the problems and issues that had been encountered, toward design resolution. Methods were proposed for simulation analysis and controller design by GE/SCSD, according to requirements and integrating contributions by G&C, particularly Mr. Bjorn M. Olson.

The initial study was devoted to the development of steady state control schedules for the Power Lever Angle (PLA) command. The PLA command to the LM-2500 translates directly to a gas generator speed command that is governor regulated, i.e., not to a fuel rate. Depending upon gas generator inlet air temperature, a considerable range of power results from a given PLA command. Emphasis was put on the need to achieve overall steady state performance without shaft speed feedback, in each mode of control, though basically for the smooth sea, full displacement and clean hull assumption corresponding to the model data. (Later studies estimated achieved ship speed degradations for effects of wind, waves and rough hull conditions.) The solution approach was to map PLA commands over the range of power and ambient temperatures for the one- and two-engine cases, and to design a method for calibration-tuning the installation to the model. The steady state data was condensed by fitting it to partitioned linear expressions. The problem was more complex for the Perry than for the SC-175, because the Perry has a more complex set of operating modes (1), including shaft idling speeds that vary as a function of ambient temperature and a quiet running mode, with restrictions on minimum shaft speed. The design proved to be effective from the beginning of the FFG-7 trials and operations, with no evident penalty to ship speed control, while eliminating the cycling of the gas generator due to shaft speed feedback and simplifying the selection of gains for the optional use speed control mode.

The final propeller test data, consisting of torque and thrust coefficient maps in relation to the unmodified advance ratio, was not available until late in the program. The early analysis was done using "representative" propeller data. The representative data could not be used for the final steady state schedules, but served well for the purposes of preliminary dynamic analyses that considered feasible control strategies. Considerable care was given to the evaluation and use of the data (G&C). It was decided at GE/SCSD that it was worthwhile to use two-way four point (third order) Lagrange interpolations, with a technique for saving indices and calculations from previous simulation cycles still useful in the current cycle. The full parameter engine simulation program had some representative load routines in it, including a ship propulsion load. This portion of the program was reworked to install the Perry propeller models and procedures. Also, a controller model was worked into the large program. At first the controller model was very simple, giving only step or ramp commands for the PLA and pitch ratio. The large program will simulate only single engine performance, or two engine performance with identical inputs and outputs, i.e., double power. The same situation was true for the SC-175 (early and late propeller data, revision of the large program). The propeller performance prediction is perhaps the weakest link in the validity of the propulsion model. Although attention was paid to the expected effects of cavitation (9), cavitation was not modeled for use in the Perry and SC-175 controller design programs. Because of the uncertainties involved, the robust nature of the control algorithms and the dynamic effects involved, the modeling of cavitation did not appear to be useful. Fortunately, the performance predictions for the FFG-7 were matched rather well in the measured trials data (1), reinforcing the opinion.

Early dynamic studies for the FFG-7 were done using the full parameter program. When the SC-175 project was started, the full parameter program was converted for use on one of a number of available Digital Equipment Company VAX 11/780 systems. Keyboard inputs with CRT display scrolling of tabulated and plotted outputs greatly expedited the simulation approach. Other major advances, including more effectively structured and transportable simulation program modules and powerful support software, are already in place for current use. The early simulations, as mentioned, were one or two engine maneuvers, with identical inputs and outputs if for two engines. The currently known methods of control, e.g., using a closed loop shaft speed rate limit, or a fixed PLA rate limit, were explored for the Perry application. Shaft speed was held at a fixed idling speed (selected separately for the one- or two-engine case) while pitch changes were executed, then pitch was held fixed for further acceleration.

In 1975 the U.S. Navy interceded in the plans of the program, to indicate that current test and operating problems had required special attention to the protection of propellers, particularly the CRP, against overthrust. Techniques such as stepping the PLA up slightly in a crash ahead until pitch reached a given level, then advancing it at a rate limit were used. Finally a nonlinear PLA command rate limit was used, while the pitch was allowed to change at the capacity of the pitch controller. The two pitch control modes resulted in either a constant pitch rate or one proportional to shaft speed (standby and attached pump modes, respectively). A large time constant was selected by simulation to assure against overthrust, as well as overtorque, over worst cases. In both the Perry and SC-175 cases overthrust without overtorque is easily done, and the thrust constraint is dominating. The reader can note the use of thrust for optimization of the SC-175 performance, as shown later (figure 3).

The requirement that DRA studies would verify by simulations that the programmed controller would assure against overthrust and overtorque without depending upon shaft speed closed loop control, or upon power cuts based on directly or indirectly measured shaft torque, led to the scheduled-adaptive design approach. The LM-2500 engine provides in its controls a calculated or predicted engine torque, based on its sensed parameters, which can be applied (along with power turbine speed and acceleration) within the engine controls to cut back the PLA command. The SC-175, then also has the signal, and in addition it has a torsion meter to directly sense shaft torque. If and when effective these measures can protect the propulsion system from overtorque, though not from overthrust. They have not proved to be accurate overall, and if set conservatively enough for all cases, are likely to prevent the attainment of full power. The Perry depends upon the predictively designed propulsion controller for torque and thrust protection.

The assurance of avoidance of overthrust and overtorque was done in relation to the worst-case maneuver, the full ahead from full astern, with two engines, and checked for other cases. A consequence of the method for nonlinear PLA command rate limiting, and a part of the cautious design approach, was that the PLA can never be commanded at higher than the full-ahead for the current ambient temperature. Another constraint, to protect the clutch in the worst-case engagement, imposed an overall PLA upramp rate limit of five degrees per second, i.e., the nonlinear rate limiter used five degrees at the low PLA command, with a decreasing rate limit at higher PLA commands. Since the PLA-to-gas generator speed relationship is nonlinear in the lower range, this allowed large increases in the ordered gas generator speed at low power levels, with lower increases allowed at higher levels. The nonlinear PLA command rate limit was not applied for astern accelerations, which used the fixed five degree rate, even with two engines on line, as this did not threaten to exceed any constraints. In order to exempt mild maneuvers from the restrictions put on the extreme maneuvers, in the crash ahead, the nonlinear PLA rate limit was imposed only when (a) the pitch is not advanced to design ahead, or (b) the shaft speed is accelerating as much as 1.8 rpm/second, or (c) the PLA command is within a set interval from the maximum allowed. Shaft acceleration is sensed using averaging of the discrete feedback inputs over a second of operations. Simulations were used to test and refine the ideas that evolved in this manner, over the cases involved. The standby hydraulic pump mode of pitch change presented a worst case, since the main pump method results in slow pitch advances at low shaft speeds.

Having established this approach for the crash ahead, it was necessary to solve the problem of crash astern in a manner compatible with its design decisions. The worst case faced was the crash astern into the low speed range from full ahead, with one engine, at the highest ambient temperature (125 degrees F), for both pitch pump modes. The overall five degree per second PLA command rate limit yields the lowest power rate at the high temperature, and the low scheduled power for the slow astern does not provide enough power to hold shaft speed to acceptable levels when the pitch goes into the negative range and the ship is still coasting ahead at a significant speed (50 percent). The approach selected assumed that the PLA would be chopped and the pitch would be ordered to its new position (after an initial drop of shaft speed, to limit spindle torque). The means selected and refined by simulation include (a) anticipation of the reversal of pitch when the positive pitch feedback reaches a given low positive value, for starting the PLA command up ramp (which is relatively negligible, to reach the scheduled level for the slow astern objective), (b) the automatic closing of the speed loop, a "forced speed" technique, and (c) dependence upon a buildup of PLA command via the shaft speed limiter function, and (d) rate limiting or freezing the pitch change when shaft speed falls below given levels. The latter measure, of course, is the most effective one for preventing excessive loss of shaft speed, because the means for building up the PLA command has a fixed rate limit. The others are effective in following through with the completion of the maneuver after the pitch rate limiting carries it through that critical phase.

Once the crash-ahead and crash-astern cases were resolved, other performance requirements were considered for controller design to compatibly satisfy other performance requirements. These included the bringing on of a second engine or taking off an engine (from programmed control); compliance with shaft braking; operations with one engine in programmed control while a second is on line in manual control; etc. The overall design and resulting performance is presented in (1), which was written and presented at the Fifth Symposium before the Perry had accumulated operational experience.

As mention by others (4), (8), it has been important not only to use full parameter nonlinear simulations to achieve available validity for predictive design, but also to develop a "simplified" model and program for expediting the volume of simulation tests needed and accommodating the peculiar interests of the design problem at hand. In the Perry program an "exercise" model was developed, using an extremely simplified model for the nonlinear response of the gas turbine. The model served the purposes of the experiments in development of the types of control techniques entertained, and critical cases were verified using the full parameter model simulation.

THE SC-175 CONTROLLER DESIGN EXPERIENCE

The SC-175 COGAG-CRP automatic controller DRA work was started in June, 1980. The approach was to adapt from the Perry design as much as proved practicable. It was recognized that the dynamic response would be considerably different with the heavier ship. The crash stop would be particularly challenging, for example, given a performance goal similar to that which had been given to the Perry (although exceeded by the FFG-7).

The early part of the DRA study was devoted to development of simulation programs for the LM-2500. The engine had been somewhat uprated for the SC-175 application, and GE/AEG had revised the performance model. The GE/SCSD plan for the DRA included the building of a new exercise model using a simplified nonlinear (SNL) LM-2500 engine simulation program model furnished by GE/AEG for its customers (10). A hybrid implementation of the model had been used by GE/AEG to furnish example simulations useful for testing other programs. The program developed is a fully digital program. The SNL model has full parameter, nonlinear representation of the PLA controller and of the main fuel controller, and capability to provide high fidelity performance prediction over adequate performance regions. The SC-175 DRA scope included the analysis and design for a separate PLA command cutback on the basis of the calculated torque signal generated in the engine PLA controller. The cutback device is analog and resides outside of the programmed control routines, and can be set to cut back the programmed control PLA command outputs, for torque protection.

An early dynamic response study was made using scaled-up propeller data from the Perry program as a "representative" propeller, approximating the performance of the planned SC-175 propeller. Simple models of the pitch controller responses to the standby and main pump modes, adapted from the Perry model, were used. The objective of the preliminary DRA was to obtain information on the nature of the crash-back problem, and on overall control as might be developed into instructions for manual control. It was assumed that the torque limiter (using torsion meter signal) would operate to protect against overtorque in the manual mode, and the operator would be required to follow rules to assure against overthrust. Various schemes involving monitoring of shaft speed, pitch, and lapses of time were considered, with interest in relatively effective schemes for good response, i.e., for use in emergency situations. It is difficult to devise really practical schemes for assuring against overthrust in manual control that do not use slow and conservative control inputs. The preliminary study made a reasonably accurate estimate of the performance (number of ship lengths) that would be available in the crash stop and the crash ahead, made a preliminary selection of the shaft idling speed (which was retained), and established the background for designing the crash stop maneuver.

It was determined by Bazan Engineering that the simplified performance model for the pitch controller of the type used in the Perry program would not be adequate for the Lips Company (Drunen, The Netherlands) equipment to be used with the SC-175. GE/SCSD was to design a forward loop for the controller, and interface it to the controls system. The main problem in the design of the forward loop was to provide positive control (quick reversal response) over a range of response variations with operating conditions. Simulations of the controller responses showed that the dead band error in the hydraulic controller main loop has components that are correlated with (a) whether the standby or main pump mode is in use, (b) the sense of change of pitch that is taking place, and (c) current shaft speed. An algorithm was designed for incorporation in programmed control to compensate the pitch commands for these conditions. Variations of oil temperature also contribute to the dead band, but the use of a mean or expected temperature was considered adequate for the compilation of the data used to develop the fit applied in the algorithms.

Separate simulations were made to determine how to rate limit the PLA commands for clutch protection. The restriction was defined in terms of a limit on power turbine acceleration at the time of engagement, for the worst case, including operating situations and ambient temperatures. It was determined that a 3.5 degree per second PLA rate limit would be required, for the standard engine, i.e., the final command is adjusted up or down for the calibration. The settling of this condition set the stage for trades on the accomplishment of the major maneuvers. The ahead acceleration, including full astern to full ahead with two engines was easily established by adaptation of the Perry approach, to protect against overtorque and overthrust. The intermediate stages of the maneuver control, not representing threats to torque or thrust limits, follow through the sequences that are applied for less extreme maneuvers, i.e., that would therefore apply for demands to start and end the maneuver at lower speeds.

The optimal control for the crash astern or crash stop is to drop thrust as quickly as possible to the limit put on astern gross thrust, then hold thrust there until the ordered astern speed is achieved, at which time a steady state would be established (11). The initial step, clearly, is to drop the PLA command sharply and start the pitch to its newly scheduled ratio at the capacity of the pitch controller, as soon as equipment conditions allow (e.g., avoidance of excess spindle torque may require a brief wait until shaft speed drops below some level). If the PLA is dropped suddenly to or near its minimum it will be more difficult to recover the shaft speed when negative thrust has been developed, or perhaps to prevent excessive negative thrust. When the PLA has been chopped to the low level, is will be necessary to anticipate the need to increase power for these reasons, and start the PLA command upward before the pitch has reversed. An alternative is to chop the PLA to some intermediate level and depend on the change of pitch to change the thrust. In any case, the approaches considered use the timely limitation of the pitch rate, based on shaft speed and shaft speed rate, to prevent loss of the shaft, and in the case of the SC-175 to prevent excessive thrust. The intermediate chop option was selected, though with retention of some anticipation of pitch reversal, to avoid an unclutch-clutch sequence. In conjunction with the intermediate chop, it was found that the pitch would have to start having its rate limited when it is reduced to about 8.5 feet. The rate limiting is achieved by reading the pitch feedback to set the pitch command back to the feedback value when it reaches the pitch mark, and subsequently incrementing the command at .3 feet per second, the value selected during the preliminary DRA.

Seven main divisions of the control functions were developed in a continuing sequence of trades, considering performance objectives and constraints, with increasing constraints to assure continuity and compatibility as the design algorithms and their switching identifiers were selected. Worst cases, applicable modes, and consideration of maneuvers within each category not needing to invoke all of the limiting required for associated worst cases, were considered in sequence. These basic acceleration and deceleration regions were developed with the assumption that other controls, as for enabling the shaft brake or recovery from braking, or for bringing an engine on or off, etc., would operate compatibly as interrupt or supplementary functions. The types of acceleration or deceleration are classified in terms of a span of ship speeds in which the ship speed demand is stopped from an upward motion (acceleration) or a downward motion (deceleration). Consideration must be given to a worst steady initial condition that the case could allow, i.e., from below or above the new command, respectively. The types are as follows:

Type 1 Acceleration - Power Down

- Invoked by raising the ship speed demand within the negative ship speed range.

- PLA is ramped down to the scheduled level without shaft speed feedback, then the operator-selected mode (speed or power) is re-established, unless the ordered speed is in the shaft idling range, in which case the speed mode is automatically invoked.

- If a pitch change is required, the change is at the capacity of the pitch controller mode in effect.

Type 2 Acceleration - Forced Speed

- Invoked by raising ship speed demand to any speed in the 0 to 50 percent range.

- While pitch is below the flat pitch region the PLA is ramped up to the level used for about 48 percent ahead steady, and is held there in the power mode.

- When pitch is above the flat pitch region the forced speed mode is invoked, to persist for a set time after pitch nears the scheduled level. Pitch is changed at the rate capacity of the mode in effect. Forced speed references a shaft speed of 70 rpm rather than the 50 rpm shaft idling speed.

- When the forced speed period is completed, the operator-selected mode (speed or power) is set, unless the ordered speed is in the shaft idling range, in which case the speed mode is set automatically.

Type 3 Acceleration - Normal Ahead

- Invoked by raising the ship speed demand to any level above the 50 percent ahead level.

- Same as Type 2, except that forced speed is used only while shaft speed is below its referenced level, and the scheduled steady state shaft speed is used as the forced speed reference level. When forced speed is no longer active, PLA up ramping is at 3.5 degrees per second for single-engine control, or for two-engine control to demands to 60 percent speed. Further PLA up ramping for two engine control is at diminishing rates as controlled by the lag function.

Type 1 Deceleration - Power Down

- Invoked by lowering ship speed demand to any level above 1/3 ahead.
- The PLA is ramped down at five degrees per second in the power mode, after which the operator-selected mode is used (speed or power), unless the speed demand is in the shaft speed idling range, in which case the speed mode is automatically invoked.
- No significant change of pitch from design ahead.

Type 2 Deceleration - Power Down with Pitch Hold

- Invoked by lowering ship speed demand into the zero to 1/3 ahead speed range.
- PLA is ramped down at five degrees per second in the power mode, after which the shaft idling speed mode is automatically invoked.
- Pitch is helo fixed while shaft speed is above 90 rpm; limited to 0.3 degrees per second when it is between 80 and 90 rpm, and allowed to change at the controller capacity when it is below 80 rpm.

Type 3 Deceleration - Forced Speed

- Invoked by lowering ship speed demand into the zero to 1/3 astern range.
- Pitch change is stopped whenever shaft speed windmills above 150 rpm.
- When not in windmill freeze, pitch moves to as low as 8.5 feet at the rate capacity of the controller mode. Below 8.5 feet it is held to 0.3 feet per second if the shaft speed is above 120 rpm, below 40 rpm, or dropped more rapidly than a set rate.
- When pitch is below 2 feet, forced speed is invoked.
- When forced speed is completed, control is in the operator-selected mode (speed or power) or automatically in the speed mode if the speed demand is in the shaft idling range.

Type 4 Deceleration - Normal Astern

- Invoked by lowering the ship speed demand to below the 1/3 astern level.
- Same as Type 3 except that forced speed is used only when shaft speed is below the forced speed reference level, and the forced speed is referenced to the scheduled shaft speed. When the forced speed is no longer active, PLA up ramping is limited to 0.4 degrees per second.

These descriptions do not show all of the details of the functions, but the additional variations are insignificant to the general explanation of the partitioning of the performance regions, and the associated selection of identifiers and control functions. Some "finishing touches" are easily added, within a performance region, using the interactive analysis procedure.

Figure 2 is a top level flow diagram of the SC-175 controller functions. The following discussion is offered to suggest the manner in which the overall controller operations are synthesized.

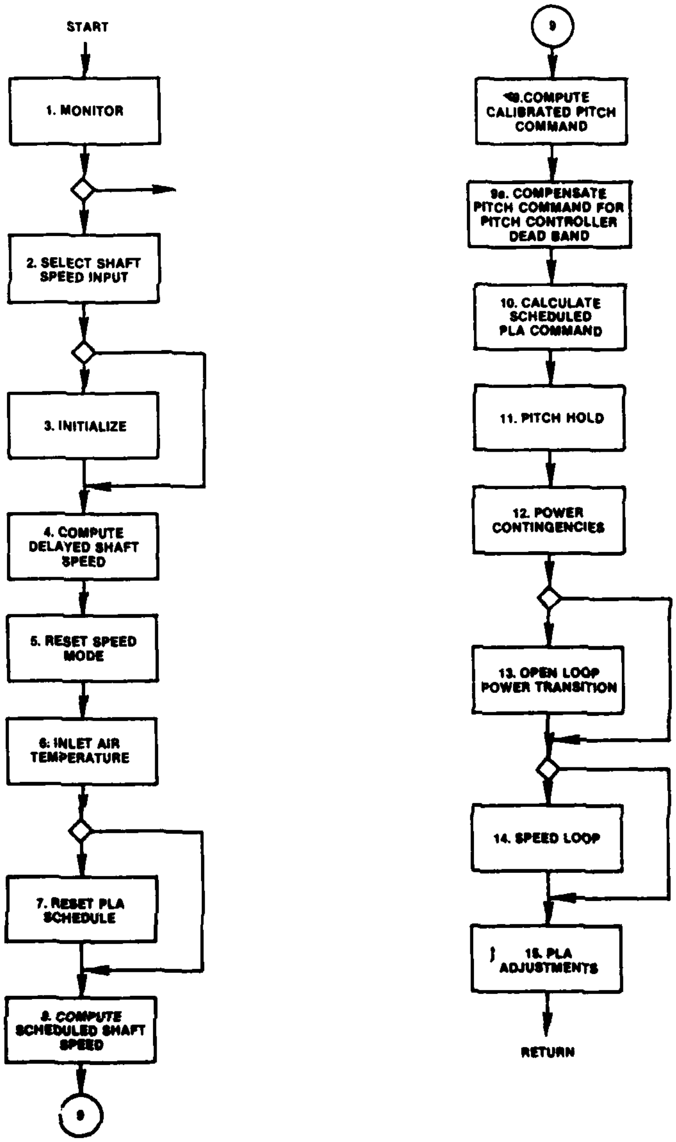


Figure 2. SC-175 Automatic Propulsion Controller Function Flow

(1) Monitor

This block is executed in every cycle of the computer used for programmed control and other functions of the more general control and logging system. The computer uses a 200 ms cycle. The monitoring block counts how many and which engines are on line and how many and which are committed to programmed control. If an engine is not in programmed control its "engine initialized" is set to "no". If no engines are in programmed control the routine is exited.

(2) Select Shaft Speed Input

If the primary shaft speed sensor input is high or low, a check is made to determine whether engine A is engaged. If it is and its speed indication is not high or low, its input is used to derive the shaft speed feedback. If it is not engaged, the same check is made for engine B. If engine B is not used, the secondary shaft speed sensor signal is used. The shaft speed will go low in some cases, as going into and out of the shaft braking sequence, and in these cases the secondary sensor will be used.

(3) Initialize

Only if no engine is initialized is this block executed. It includes 17 statements to set values that will trigger algorithms for initialization when they are first invoked. These sets are maintained downstream by resets for initialization of a given routine each time that the control flow decides to bypass the execution of that routine.

(4) Compute Delayed Shaft Speed

The selected shaft speed feedback signal is put through a lag function, generating a smoothed, delayed shaft speed input. The smoothing serves to reduce noise in the discrete input, from line voltage variations prior to the analog-to-digital converter input to the processor. The delay is used for comparison with current inputs to estimate shaft speed rate (some noise on the input for current shaft speed is tolerable in the process). The choice of lag, time constant for the rate sensing objective is established empirically by simulation.

(5) Reset Speed Mode

The operator can switch between the speed and power mode, but programmed control may again switch speed on or off, depending upon circumstances. This routine determines whether the braking procedure has been enabled, or is on, and if so, sets a flag to that effect for downstream use, which will route the flow to assure that the speed loop is opened. If the braking is not enabled or on, the flag is set oppositely, and a check is made to determine whether two engines are on line while only one is in programmed control. If so, and if the engine not in programmed control is engaged, the speed mode is set to on and the "parallel control" flag is set to on. If the parallel control flag is not set to on, it is set to off, and a check is made to determine whether the demanded ship speed is in the shaft idling range. If it is, the speed mode is set to on.

This is not the only possible reset of the speed mode. Forced speed invocations set it on, and PLA ramping to newly scheduled levels set it off, downstream.

(6) Inlet Air Temperature

The inlet air temperature (compressor inlet) of each engine in programmed control is read, and if two engines are in programmed control mode, the lower air temperature is selected for control calculations. The temperature is limited between set values. If a difference in temperatures is to be used, a downstream calculation will calculate the partial differential of PLA with respect to temperature, for final adjustment of command to the engine having higher inlet temperature.

(7) Reset PLA Schedule

If initializing the scheduling, or if a change of the number of engines has been made since the last computer cycle, this routine shifts an initial trial index to be within the range of the one or two engine arrays. Minimum and maximum accepted ship speed levels are set.

(8) Compute Scheduled Shaft Speed

The previously set index for the piecewise-linear shaft speed curve segments is bumped up or down when the demanded shaft speed passes above or below its ship speed limits. When the index is fixed, the scheduled shaft speed is calculated as a linear function of the ship speed demand, using coefficients for the indexed interval.

(9) Compute Calibrated Pitch Command

Since a fixed shaft idling speed is used for both one and two engine applications in the SC-175, the scheduled pitch is calculated in the same way as explained for scheduled shaft speed. The routine then determines whether the pitch command, for the test propeller, is on the interval extending from and below the command for zero speed, or from and above it. It selects two pitch calibration inputs (for neutral and design astern, or for neutral and design ahead), and prorates the bias to be added to the command.

(9a) Compensate Pitch Command for Pitch Controller Dead Band

This routine calculates the dead band expected error in terms of whether (a) the standby or main pump is in use, (b) the sense of the pitch change commanded, and (c) the current shaft speed. Seven different linear functions of the calibrated pitch command are used for the cases distinguished, and the error is integrated with the command to compensate the command so as to cancel the error.

(10) Calculate PLA Command

The PLA scheduled command is calculated as a linear function of ambient air temperature, but the two coefficients used are first calculated as linear functions of the portion of the ship speed interval indexed. The partial derivative of FLA with respect to ambient temperature is saved from the calculations. The maximum PLA command allowed for the conditions at hand (four, including one or two engines, with ahead or astern propulsion) is computed.

(11) Pitch Hold

This routine contains the logic for rate limiting or freezing of the propeller pitch.

(12) Power Contingencies

This routine implements the PLA commands for the major types of accelerations through the release from forced speed.

(13) Open Loop Power Transitions

If forced speed is not active or the prereversing power step used when thrust reversal is anticipated is not on, or if the shaft braking procedure is not being accommodated, this routine is entered. If it has not previously turned on its algorithms and has not detected their objectives satisfied, it tests whether to turn them on because (a) a significant change of ship speed demand has been made, (b) a change of the number of engines in programmed control has been made, or (c) a change from the speed to the power mode has been made. In any of these cases its algorithms will compute appropriate incrementations of the PLA command for open loop control until the current PLA commands correspond to the current scheduled commands, however the scheduled commands may vary during the procedure.

If two engines are in programmed control and their PLA's are not equal and not at the scheduled level (the general case for bringing a second engine into programmed control), the outer command is brought to the inner, then the two together are brought to the scheduled level. Down-ramps are at five degrees per second, but *upramps must go through the upramp algorithm.* If the two PLA's are found to bracket the scheduled level, they are ramped linearly at rates so they will arrive at the scheduled level at the same time (for fixed scheduled level, else reset this way when the scheduled level is changed).

The upramp routine incorporates the various limits and conditions described for the major acceleration categories.

(14) Speed Loop

This module is executed when the shaft speed feedback loop is closed, or dumps the accumulated integral term in the proportional-integral error compensation and exits. The measured shaft speed error is limited in magnitude, and if the "heavy seas damping" mode has been switched on, it is reduced. Special gains are set for the forced speed or parallel control conditions, or it is reduced inversely in proportion to the scheduled PLA level in ratio to the maximum PLA command. Adjustable inputs for a range of gains for the proportional and integral terms in the compensation are read and incorporated. If shaft speed is above a set level, a limit is put on the PLA command overshoot, to prevent overtorque during turns in the speed mode. The integral term is trimmed at the level which yields the maximum or minimum PLA command, to improve positive control. A two stage scaling technique is used for the integer arithmetic, so that slow command changes can result from small shaft speed errors (without loss of effect because of integer truncations).

(15) PLA Adjustments

This routine includes the invoking of the accumulation of a command correction term for the PLA if shaft speed goes above or below set limits, with gradual degradation of the term toward zero when it has gained a value and the shaft speed is within acceptable bounds. This procedure is reconciled with the shaft braking operations, including the monitoring of the increase of engine speed when shaft and turbine speeds are very low, in which case engine speed is limited. Finally, PLA calibrations bias is allocated to the outgoing PLA command.

Figure 3 shows a performance prediction for a crash astern in one engine control. Ship speed is ordered to go from full ahead (84 percent) to full astern. The curves are plotted in terms of percentages, e.g., pitch starts at design ahead or 100 percent. The PLA command is chopped immediately to an intermediate level. Shaft speed shows some windmilling, after an initial drop. Pitch is allowed to drop at the capacity of its controller until it has reached about 8.5 feet, at which time its commands limit it to the slow rate of change. Gross thrust has been dropping very rapidly, and now slows its drop rate, and stops dropping when it reaches its allowed limit. When pitch reaches two feet, the PLA command is allowed to increase toward the scheduled, full astern level, but at a very slow rate, to continue to avoid overthrust. As the shaft speed continues to drop and pitch continues to change, gross thrust stays in the neighborhood of its limit. The two engine crash astern is similar, but shows a brief period of pitch freeze due to excessive windmilling, because shaft speed and ship speed starts at a higher level.

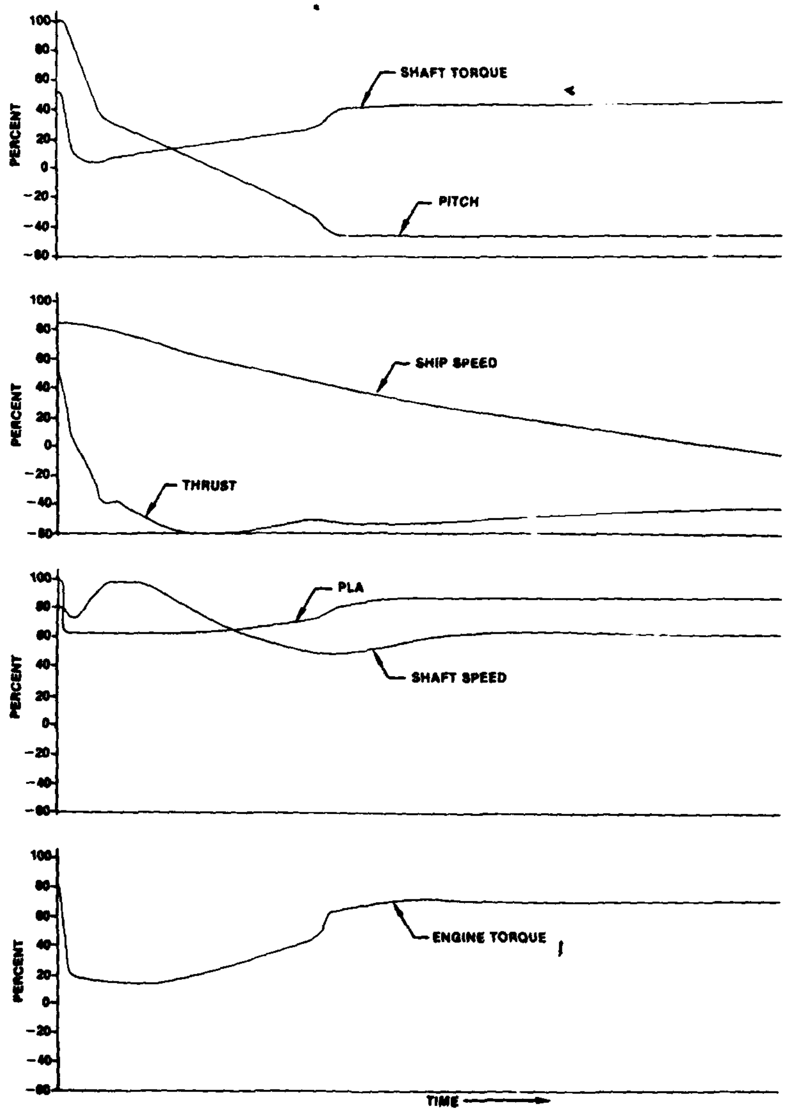


Figure 3. Performance Prediction - Crash Astern Maneuver, One Engine

Figure 4 shows a prediction for a full-astern-to-full-ahead maneuver with two engines. Gross thrust dwells at its allowed upper limit, in this extreme case, while propeller torque is relatively moderate. Figure 5 shows a crash ahead from dead in the water with one engine. Note that engine torque tends to dwell near its allowed limit, in this optimization, while propeller torque and thrust are relatively moderate.

CONTINUED CONTROL STUDY: THRUST LIMITING CONTROL

The LM2500 engine control is designed to protect the propulsion system from shaft overspeed and overtorque conditions. Ship propulsion dynamic response analysis indicates that during transient maneuvers the propeller can overthrust without overtorque. As was pointed out in the previous section, the propulsion programmed control used in the Perry class frigate and the Spanish Carrier SC175 takes care of the overthrust, overtorque, and overspeed condition based on predicted ship dynamic response analysis.

An SCSD internal research and development study was successfully carried out to develop an on-line thrust cutback controller (12, 13). The controller consists of two steps, first the propeller thrust is estimated using sensed propeller torque, shaft speed, and propeller pitch as inputs; then the current torque coefficient is computed, and the propeller characteristics maps are interpolated to compute the advance ratio. Consequently the propeller thrust can be estimated. In the next step the error between the estimated thrust and the limit propeller thrust valve is used to cut back fuel using the engine PLA.

Several extreme cases were simulated in manual control using the Spanish Carrier Simulation Program. These cases indicated that with the thrust fuel cutback controller the propeller can be protected during extreme transient maneuvers in manual control mode.

CONCLUSIONS

Because of its pragmatic nature, the scheduled-adaptive approach especially depends upon a special network of individuals to actively contribute to the analysis and design. This has been put well by Tiano et. al. (4) in relation to the development of adaptive controls for steering and course-keeping. While the digital simulation capabilities of the computer and the flexible implementation capabilities of the onboard processor have been the enabling elements for the controls, the concepts and assessments used in developing the controller must come from the very best and deepest understanding and scope of experience; i.e., personal knowledge and judgment of those dedicated in areas of operations, ships machinery, the gas turbine design, simulation and test, CRP design, etc., depending upon the case in hand.

The use of simulated time rather than real-time simulations provides the DRA development with greater flexibility in the use of available, high confidence models and modules. A great leverage is gained by using generally available general purpose computers and associated support software. The program editing and interactive display capabilities of the smart terminal, and the increasing availability of skills associated with these is very favorable to the needs. Similar advantages are available in the general development of better methods for structuring, modularizing, commenting, etc. Program modes for emulation of the digital controller's integer arithmetic (if applicable) and for preparing and formatting data to simplify and ease the development of machine language coding (if applicable) and its testing, and to similarly expedite real-time tests with system hardware interactions, are readily available. The simulated time, higher language approach will also serve the development and maintenance of real-time simulators as appropriate, while perhaps making it more attractive to do some real-time tests that exploit the cabling and interconnecting work supported by the application system engineering program, as such. The DRA can contribute to design integration and automation. The DRA programs can be planned to be transportable to field and user installations, and packaged to include a software description document and a user's manual. The continued maintenance of the DRA program in relation to ship and fleet experience will not only enable better entertainment and execution of change and growth developments, but will provide invaluable feedback of knowledge to support the development of better designs.

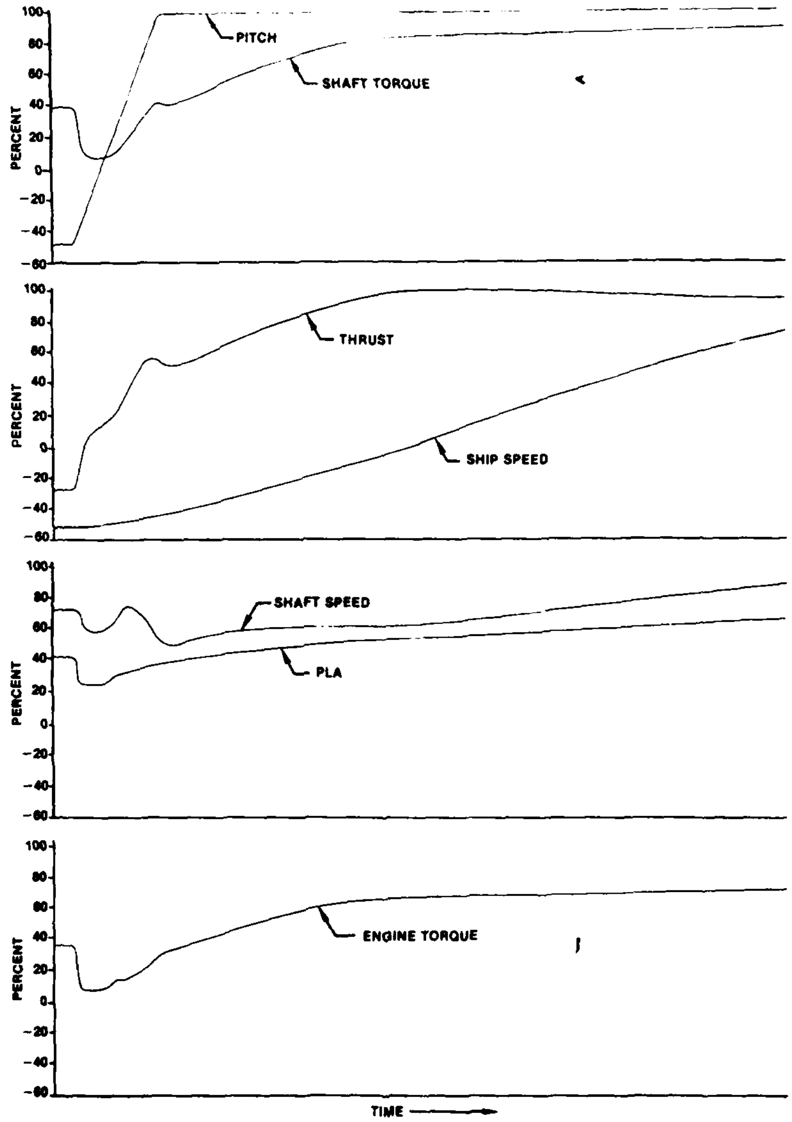


Figure 4. Performance Prediction - Full Astern to Full-Ahead Maneuver, Two Engines

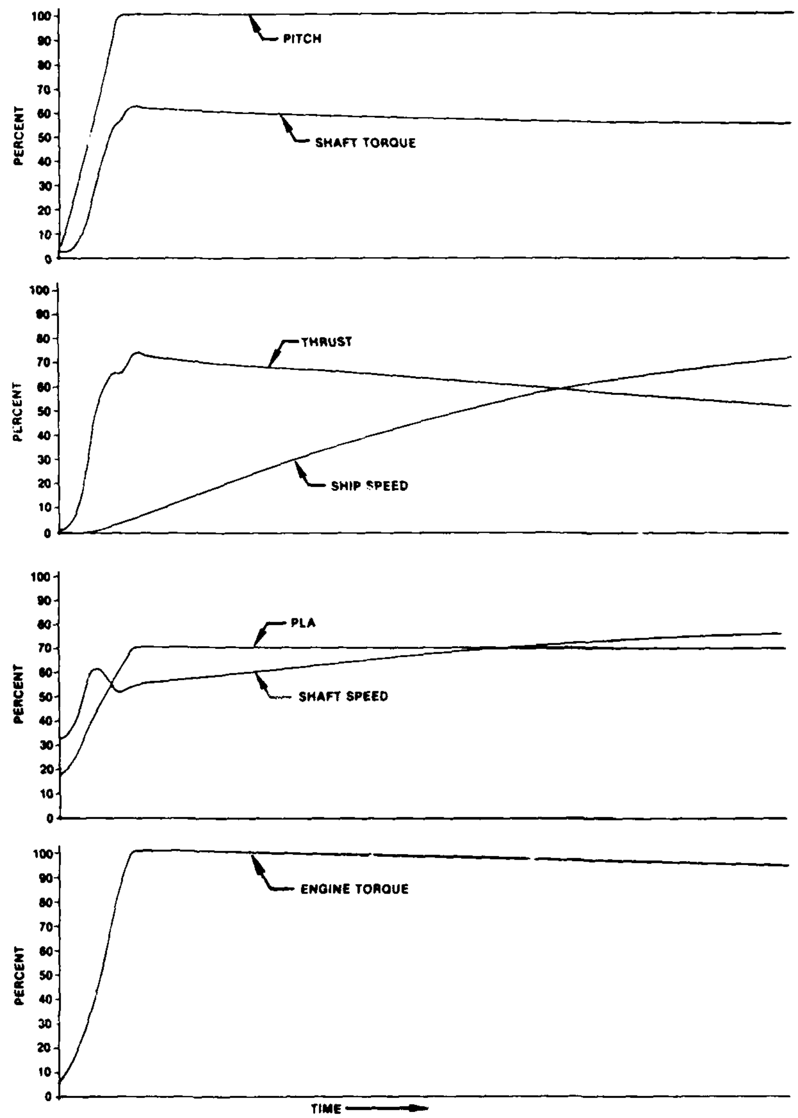


Figure 5. Performance Prediction - Crash Ahead from Dead in the Water, One Engine

ACKNOWLEDGMENTS

The authors wish to acknowledge special contributions of those who participated in and guided the progress of the Perry and SC-175 controller design studies. M. Hartley coded and operated the SC-175 simulations and coded and verified the machine language software for the onboard processor. P. Gamble similarly served the Perry program. B. Desai supervised the Gibbs & Cox Company supported of Bazan systems engineering. B. Olson of the Gibbs & Cox Company contributed substantially to the development of the Perry and SC-175 controller concepts, providing detailed functional requirements, ships machinery interface data and integration, a wealth of judgments on practical considerations, and systematic reviews of the analytical work. C. Norris provided Program Engineering leadership. D. Moss and E. Robinson provided broad systems engineering and integration for the controllers, within the total control systems provided by General Electric. A. Wesselink of the Lips Company, Netherlands, provided detailed data, direction and assessment for the development of the pitch controller simulation. Representing the General Electric engine manufacturers, S. Smith expedited integration of design data and P. Mingler provided expert guidance and assessment on simulation of the LM2500 performance. A. Nitsch organized and managed the development and editing of this report.

REFERENCES

(1) C.W. Cassel, M. M. Olson and A. E. Dresser, "FFG-7 Class Propulsion Controls - Design and Dynamic Performance", Proceedings of the Fifth Ship Control Systems Symposium, October 30 - November 3, 1978, at the David W. Taylor Naval Ship Research and Development Center, Annapolis, Maryland, USA, Volume 5.

(2) R. H. Purnell, Commander, U.S. Naval Reserve, "New Carrier for Tomorow's Spanish Armada", Proceedings of the U.S. Naval Institute, March 1984.

(3) P. C. Parks, W. Schaufelberger, Chr. Schmid, H. Ungehauen, "Applications of Adaptive Control Systems", Proceedings of an International Symposium at Bochum, May 1980, "Methods and Applications in Adaptive Control", edited by H. Ungehauen and published as Volume 24, "Lecture Notes in Control and Information Sciences", edited by A. V. Balakrishnan and M. Thoma, published by Springer-Verlag, Berlin-Heidelberg-New York.

(4) A. Tiano, E. Volta, A. W. Brink, "Comparison of Various Adaptive Control Techniques Applied to Autopilot Design", Proceedings, Sixth Ship Control Systems Symposium, 26-30 October 1981, Director General Maritime Engineering and Maintenance National Defence Headquarters, Ottawa, Ontario, Canada, Volume 4.

(5) Ya. Z. Tsypkin, "Adaptation, Training and Self-organization in Automatic Systems", Automatic Remote Control, Volume 27, 1966.

(6) B. K. Tanner, A. M. Dorrian, D. W. Anderson, K. W. McTavish, "Simulation as a Tool in Warship Control System Design and Development", Sixth Ship Control Systems Symposium, 26-30 October 1981, Director General Maritime Engineering and Maintenance National Defence Headquarters, Ottawa, Ontario, Canada, Volume 1.

(7) P. T. Kidd, N. Munro, D. E. Winterbone, "Multivariable Control of a Ship Propulsion System", Sixth Ship Control Systems Symposium, 26-30 October 1981, Director General Maritime Engineering and Maintenance National Defence Headquarters, Ottawa, Ontario, Canada, Volume 4.

(8) C. J. Rubis, R. R. Peterson, "Simulation LM 2500 Marine Gas Turbine Dynamics", Third Ship Control Systems Symposium, Bath, England, September 26-28, 1972.

(9) C. J. Rubis, "Acceleration and Steady-State Propulsion Dynamics of a Gas Turbine Ship with Controllable-Pitch Propeller", The Annual Meeting of the Society of Naval Architects and Marine Engineers, New York, N.Y., November 16-17, 1972.

(10) "LM2500 Simplified Nonlinear Engine and Control System Simulation", MID-TD-2500-13, Revised August 1978, Marine and Industrial Projects Department, General Electric Company, Cincinnati, Ohio, USA.

(11) J. W. Donneley, "Propulsion Control Optimization of CRP Propeller Driven Ships", Fifth Ship Control Systems Symposium, Oct. 30 - Nov. 3, 1978, David W. Taylor Naval Ship Research and Development Center, Annapolis, Maryland, USA, Volume 5.

(12) C. W. Cassel, "An Algorithm for Computing Propeller Thrust and Ship Speed as a Function of Propeller Torque, Pitch Ratio, and Shaft Speed", IR&D Task 682661, 12 Aug. 1982.

(13) E. Pollak, "Advanced Control Concepts", GE-SCSD Report PIR 661-IRD-8301-001, December 29, 1982.

CHARLES W. CASSEL BIOGRAPHY

Short biography for Charles W. Cassel, principal author of "Optimal Design of Digital Propulsion Control for the Spanish Carrier SC175: Associated Trends", Session 31.

Charles W. Cassel is a Senior Systems Engineer with the Simulation and Control Systems Department of the General Electric Company, Daytona Beach, Florida, USA, where he designed the digital propulsion control systems for the U.S. Oliver Hazard Perry Class Frigate and the Spanish Carrier SC175. He graduated with an AB in Math at Wittenberg University, Ohio, and later pursued graduate programs in math and systems analysis at the Ohio State University and at the University of Florida. His current responsibilities include the design of software systems and special algorithms for the development and implementation of vehicle systems training simulators.

EYTAN POLLAK BIOGRAPHY

Dr. Eytan Pollak received his BSAE and MSME from Technion-IIT, Haifa, Israel, in 1974 and 1976 respectively, and a PHD ME from Purdue University in 1978.

He is currently a Control Systems Engineer with primary responsibilities for Control System Design and Analysis, as well as the development and implementation of systems simulations. He has applied his efforts within the fields of propulsion systems for ships and locomotives as well as ship's electric plants. He has worked on numerous systems using gas turbine, diesel, steam turbine, and dc motor prime movers, as well as predictive servo controls for advanced electro-optic systems.

His previous efforts concerned design and simulation of compressor and heat pump capacity controls. Prior to his General Electric experience he held research positions at Purdue and Technion, working on gas compressors and energy conservation systems. He started his career in the Israeli Navy as a system technician for ships machinery and auxiliary systems.

Dr. Pollak has been published in the fields of mechanical and electrical systems simulation, control, and optimization.

CANADIAN PATROL FRIGATE MACHINERY CONTROL SYSTEM

by Rashid Khan
and Peter Eich
CAE Electronics Ltd.

ABSTRACT

CAE Electronics Ltd. has been selected to provide a computer based Integrated Machinery Control System (IMCS) for the Canadian Patrol Frigate (CPF) Program. This paper describes the system that will be implemented with particular emphasis on the control system concepts and the top-down modular approach to software design.

INTRODUCTION

The Integrated Machinery Control System (IMCS) for the Canadian Patrol Frigate (CPF) is based on the Canadian Department of National Defence Shipboard Integrated Machinery Control System (SHINMACS*). SHINMACS is the result of considerable study and development to determine the machinery control concepts necessary for a modern warship such as the Canadian Patrol Frigate.

These concepts, together with the availability of a number of building blocks called Standard Digital Equipment (SDE), provide a high level of confidence that the proposed system will not only meet the specified machinery control system requirements but will also offer a number of other benefits. Typical of these benefits will be the reduced life-cycle costs resulting from the use of SDE for other systems in the ship and the attendant savings in areas such as documentation, training and spares.

One of the fundamental objectives of SHINMACS is to substantially improve the man-machine interface. To this end, the Defence and Civil Institute of Environmental Medicine (DCIEM) has carried out extensive investigation and analysis resulting in the publication of a three-part report titled Human Engineering Design Requirements for SHINMACS Machinery Control Consoles (1). The proposed machinery control consoles incorporate the DCIEM recommendations. CAE Electronics Ltd has been designated as the supplier of a SHINMACS Advanced Development Model (ADM). The work to be carried out under the ADM contract will result in a system where the concepts and existing building blocks will be brought together in a single operating system.

CAE has extensive experience as a supplier of real-time computer systems covering a wide range of diverse applications, from the systems which monitor and control nuclear power plants, to flight simulation for a host of military and commercial aircraft.

The combination of this diversified experience coupled with the extensive SHINMACS development already carried out and the work pro-

* Trademark of the Department of National Defence (Canada)

ceeding under the SHINMACS ADM provides the high level of confidence necessary to supply a system on schedule which will meet the operational requirements of the Canadian Forces.

CONTROL SYSTEM CONCEPTS

The Shipboard Integrated Machinery Control System (SHINMACS) and the Shipboard Integrated Processing And Display System (SHINPADS*), which is part of SHINMACS, have been the subject of papers presented in previous symposia. However, in order to aid in the understanding of the proposed IMCS configuration for the CPF, a brief description of each concept is given below:

SHINMACS is a Canadian Forces machinery control concept which will be used to implement current and future control systems for propulsion, ancillary and auxiliary plants. SHINMACS introduces a step forward in the monitoring and control of shipboard plant, by combining the reliability and flexibility of digital computers with the simplicity and the ergonomics of an advanced graphics display based man-machine interface.

The system is based on a distributed architecture, and utilizes redundant serial data buses to link the supervisory computers at the operator consoles in the machinery control space with the microprocessor based data collection units in the machinery spaces. The reliability and maintainability of the system is enhanced by its military-qualified, digital electronics design, while the use of software and firmware for implementation of the data collection and control algorithms ensures flexibility for system tuning and for ease of modification to meet future requirements.

The primary man-machine interface to the system utilizes computer generated colour graphics and text displays to present plant monitoring data. Presentation in this way ensures that significant items, such as alarms, can be highlighted in order to attract the operator's attention, while routine data not relevant to the task in hand is not permitted to distract him, but can be held in the computer ready for presentation when he is ready for it. Naturally the operator is made aware of an alert condition on any system as soon as the event occurs. Control of the shipboard plant is achieved by the use of functional keyboards. These provide some dedicated push buttons for special purposes such as engine trip, as well as 'soft' keys whose function is determined and indicated by the control sequence in progress. By limiting the use of these keys to functions which are valid for the current state of the plant and the point reached in the control sequence, the possibilities for human error are significantly reduced. Schematic presentations are generally used for both monitoring and control data, in order to ensure that information appears in an easily comprehensible form.

The data processing power provided by the digital computers permits the operator to use the concept of system control, rather than that of unit manipulation. Thus while the system supports manual control of gas turbines and propeller pitch, it also accepts commands for ship speed; in the latter (automatic propulsion) mode, shaft RPM and propeller pitch are determined automatically from demand schedules.

* Trademark of the Department of National Defence (Canada)

The ergonomically designed integrated control consoles are constructed for minimal size and weight. This design objective complements the weight and installation effort savings realized by the significant reductions in cabling achieved by the use of the distributed system architecture. Transducers need only be cabled to the closest data collection unit; data will then be transmitted over the serial data buses to other points in the system where it may be needed.

SHINPADS is a complete interprocessor communication system for a realistic and reconfigurable interconnection of micro-processors. The Serial Data Bus (SDB) includes the Bus Transmission System, bus interface devices called nodes and real-time system software necessary for implementing distributed computing networks. Its multiple-redundant data paths, 10 MBPS bandwidth, and standard interface provide the high reliability and maintainability necessary for demanding command and control system architectures.

SYSTEM HARDWARE

A simplified version of the IMCS system architecture is illustrated in figure 1. It can be seen that the configuration is modeled on a SHINMACS design. Control console computers interface to dual-redundant digital controllers through the high-speed AN/UYC-501 SHINPADS serial data bus, while the intelligent Remote Terminal Units (RTU) communicate with the digital propulsion controllers through a dedicated highspeed data acquisition bus (DACBUS-M).

The Bridge, Machinery Control, Maintainer/Electrical and Supervisory Consoles each have local data processing capability and are fitted with advanced man-machine interface (MMI) facilities utilizing video display units with computer generated graphics and programmable function push buttons.

Man-Machine Interface Group

The man-machine interface group will provide the active control devices and colour displays to permit the operators to control, respond to or interrogate propulsion, ancillary and auxiliary systems. Information will be presented to the operators in the form of CRT pages. The CRT pages will act as windows into a large array of data which may include parameter readings in digital or graphic form, messages and schematic system diagrams which can be overlaid with data updated in real-time. A typical SHINMACS ADM CRT page is depicted in Figure 2.

The above functions will be provided at each MMI console, however the operator control capability at each console will vary according to the requirements of the station. Detailed below are the functions, capabilities and configuration provided at each MMI station.

Bridge Control Console (BCC). The BCC will provide MMI facilities to support control and monitoring functions for bridge personnel. Data processing capability will be provided by the AN/UYK-502 Computer. The MMI features include a CRT display together with functional keys and audible alarm indicators.

The CRT display will provide all the necessary information required for automatic control.

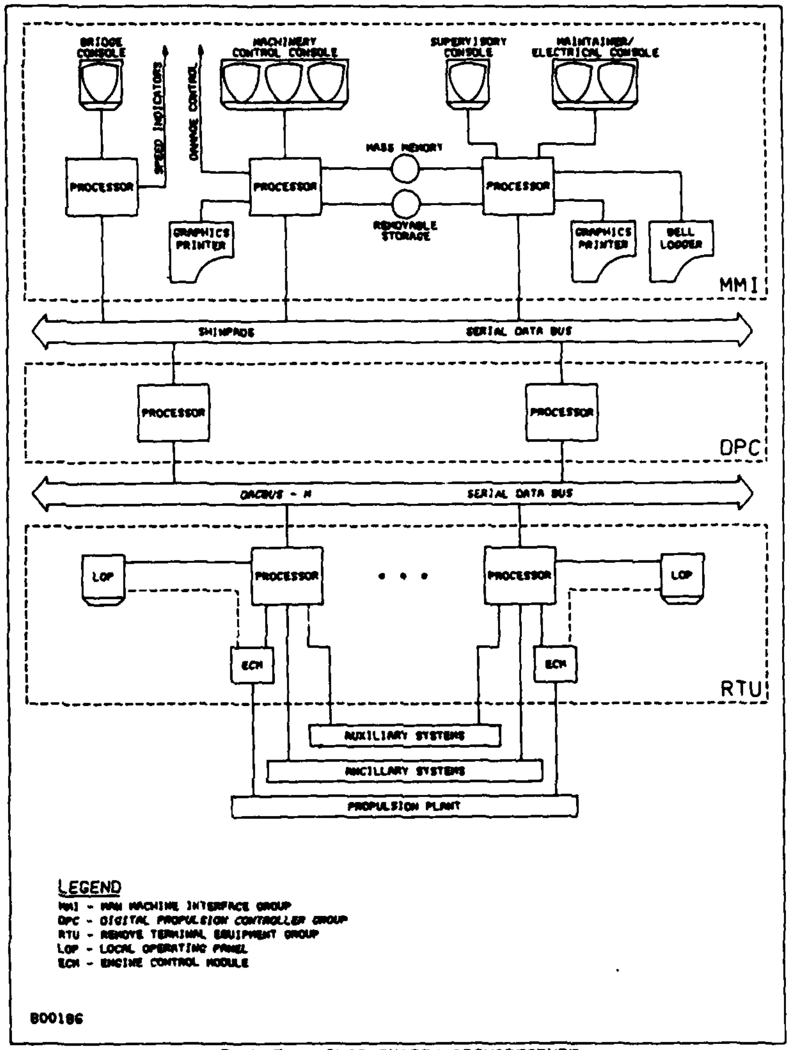


FIGURE 1. IMCS SYSTEM ARCHITECTURE

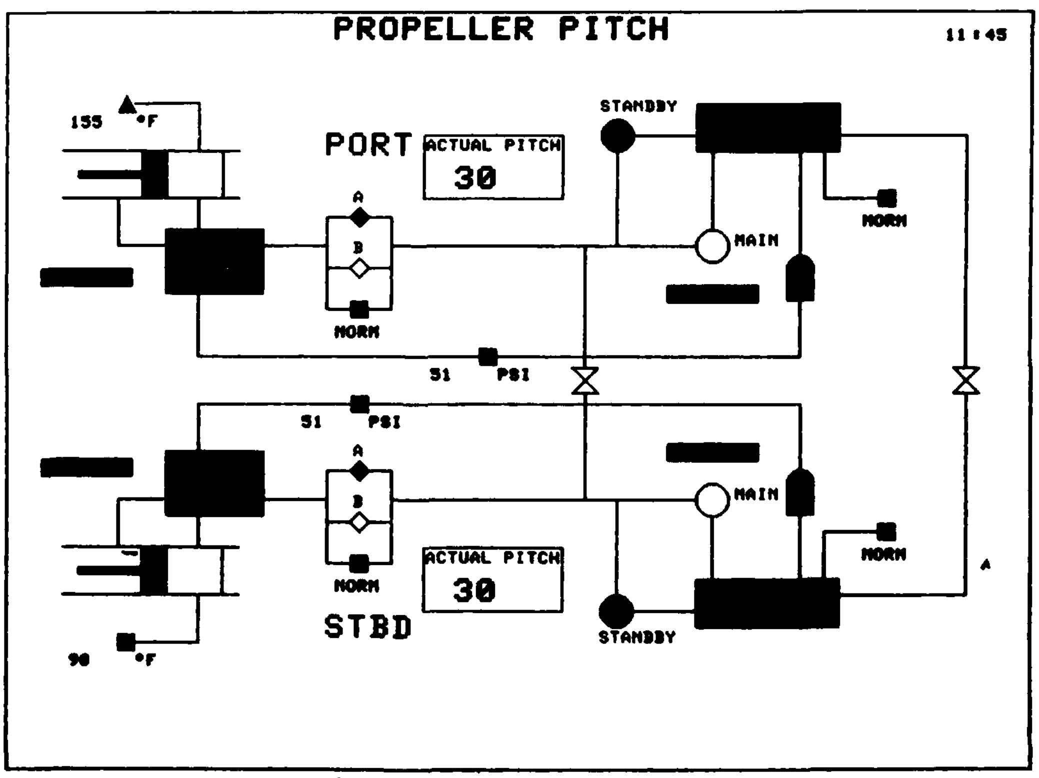


Figure 2, A Typical SHINMACS ADM CRT Page

Machinery Control Console (MCC). The MCC provides the main MMI facilities enabling the operator to control the propulsion plant including ancillary and auxiliary machinery.

Data processing capability will be provided by an AN/UYK-502 computer. The MMI features include three CRT displays, function keys, audible alarm indicators and keylocks.

Automatic and manual control of the engines and propeller pitch will be carried out using four port joysticks and four starboard joysticks.

In the automatic mode, the knots joysticks are used to enter ship speed. The shaft RPM and propeller pitch required to generate the requested speed are derived from the demand schedules and output to the plant. In the manual mode the engine speed control and propeller pitch joysticks are used to enter the shaft RPM and the propeller pitch setting to achieve a desired ship speed.

Supervisory Control Console (SCC). The SCC's primary function will be to provide MMI facilities for the control and display of data required for supervisor monitoring of the machinery and control system. This will include the display of data and the selection of parameters for the compilation and display of data from logs, dynamic analysis, and the health and performance monitoring systems. The SCC's secondary function will be that of a backup position to the MCC in case that the MCC was inoperable. To provide this dual redundancy of operating positions in the Machinery Control Room, the supervisory console must provide the MMI facilities to perform all of the functions that are available at the MCC; this will include the control and monitoring of the propulsion machinery, ancillary and auxiliary machinery, and the selection of propulsion control mode and station-in-control.

Data processing capability will be provided by an AN/UYK-502 computer. The MMI features include a CRT display, function keys and audible alarm indicators. A numeric keypad entry module will also be provided to input data required for the health monitoring functions.

Maintainer's/Electrical Control Console. This console provides the facilities to monitor and control the electrical generators and distribution system. The second CRT and keypad is used to perform IMCS diagnostics and maintenance functions.

Data processing capability will be provided by an AN/UYK-502 computer. The MMI features include 2 CRT displays, functional keys and audible alarm indicators.

Bus Equipment Group

The SHINPADS and DACBUS-M SDBs will provide the communication mediums for the IMCS. Both of these buses will provide multiple redundant paths interconnecting intelligent devices that are distributed throughout the ship, therefore increasing the reliability and combat survivability of the machinery control system and facilitating future changes to the system.

Digital Propulsion Controller Equipment Group

Both the Digital Propulsion Controllers (DPC) will be AN/UYK-502 micro-computers. The DPCs will be responsible for the control and monitoring of the propulsion plant.

Each DPC will have the capability of monitoring and controlling both shafts. The two DPCs will operate in a dual redundant mode. One DPC will act as the 'master' DPC, and the other DPC will act as the 'standby'.

Both DPCs will continuously monitor the plant. Only the 'master' DPC will output control commands to the plant. In the event of the failure of the 'master' DPC, the 'standby' will assume control of the plant.

Local Operating Panels

Two local operating panels (LOPs) are provided. One for the aft engine room and one for the forward engine room. These panels will serve as last-resort emergency control centers, should the computer system suffer an outage. The LOPs will carry hard-wired instruments and controls that will be suitable for emergency manual operation of the propulsion machinery. The LOPs will also allow limited automatic controls to be initiated manually but carried out automatically by the DPCs.

Engine Control Module (ECM)

A particular requirement of the project is the development of the ECM. This unit replaces the LM2500 turbine FSEE (Free Standing Engine Enclosure). Forming the interface between a gas turbine and the RTU group, it contains hard-wired circuits which provide essential functions for engine operation and protection. These functions include the PLA control electronics, basic start and stop sequencing, acceleration/speed and torque limiting, overspeed shutdown, etc., as well as associated built-in test capabilities.

Remote Terminal Units (RTU's)

The RTU will receive and transmit messages to the dual-redundant digital propulsion controllers through ports conforming to the EIA STD-RS422 standard, which provides for high immunity to electromagnetic interference (EMI). The communication protocol used will be the Advanced Data Communication Control Procedures (ADCCP) protocol. The protocol provides high message security through the use of a 16 bit cyclic redundancy check code (CRC-16) and flexibility and efficiency through the use of powerful addressing and message structures.

The modules used in the RTU will consist of the following three groups.

i) The Datapath Microprocessor Controller (DMC). The DMC is an INTEL iAPX 186/10 based microprocessor.

ii) The communication and control unit which provides the ability to communicate through 3 RS422 links with the DPCs.

iii) Process I/O - digital and analog input and output modules for interfacing with the plant transducers.

SYSTEM SOFTWARE

In order to produce software which is more readily understood, easier to debug and inherently more reliable CAE enforces the design concepts and considerations described below.

Stepwise Refinement - A Top Down Design Technique. In this technique the solution to a complex problem is first expressed using general statements. Each statement is then expanded or refined into more detailed statements. This process continues until all the statements have been expressed in terms of a programming language.

Every refinement step implies design decisions. It is important for the programmer to be aware of the underlying design goals in order to make the 'correct' decisions.

Structured Programming. Structured programming is a technique which uses a limited number of "logical constructs" that tend to minimize the complexity of program flow and keep each element of a program manageably small.

Structured programming is built on three logical constructs (with some permutations):

- "Sequence" - The execution of one task followed immediately by another task.
- "IF-THEN-ELSE" - The execution of a "THEN-task" when a decision is true, or alternatively, the execution of an "ELSE-task" if the decision is false.
- "Repetition" - A task that is executed repetitively until a predefined condition is met.

Programming Languages

High-level languages lend themselves to structured programming far better than assemblers and, for this reason, the majority of the software for the IMCS will be coded in a high-level language. Only time-critical code segments such as the interrupt service routines of device drivers will use the assembly language. The high level languages available on the AN/UYK-502 are FORTRAN and CMS-2/16. The latter has been chosen because it lends itself to structured programming and structured data definitions.

The high level language selected for the Remote Terminal Unit is PL/M. This is a structured programming language and is the standard high level language used for firmware by CAE.

Software Development

The software and software documentation for the IMCS will be developed in accordance with the Director General Maritime Engineering and Maintenance (DGMEM) Naval Software Standards.

Figure 3 illustrates the design process that will be followed in the development. This design process has already been applied to the SHINMACS ADM software development.

The design process illustrated in Figure 3 allows the development of the software beginning from the top down. Information determined from the requirements, becomes a tool that leads to an overall representation of the software. In addition, the preliminary design stage results in the definition of the interface among internal software elements and external system elements.

In order to meet the requirements of IMCS the software will be organized into the following groups:

1. Man-machine Interface Software.
2. Digital Propulsion Controller Software.
3. Remote Terminal Unit Firmware.

Man-Machine Interface

The software used to provide the interface between man and machine is known as man-machine software.

The primary functions of this software are to accept the operator inputs, and either provide the displays required for the monitoring and control functions or output resulting commands to the DPCs to accomplish plant control. The secondary functions are to provide data logging and the machinery health monitoring.

The health monitoring functions will include:

- Trend Analysis,
- Vibration Analysis,
- Gas Path Analysis,
- Dynamic Analysis,
- Archiving.

Digital Propulsion Controller Software

The Digital Propulsion Controller will provide a suite of software to acquire data from the Remote Terminal Units and to perform control of the propulsion, ancillary and auxiliary plant.

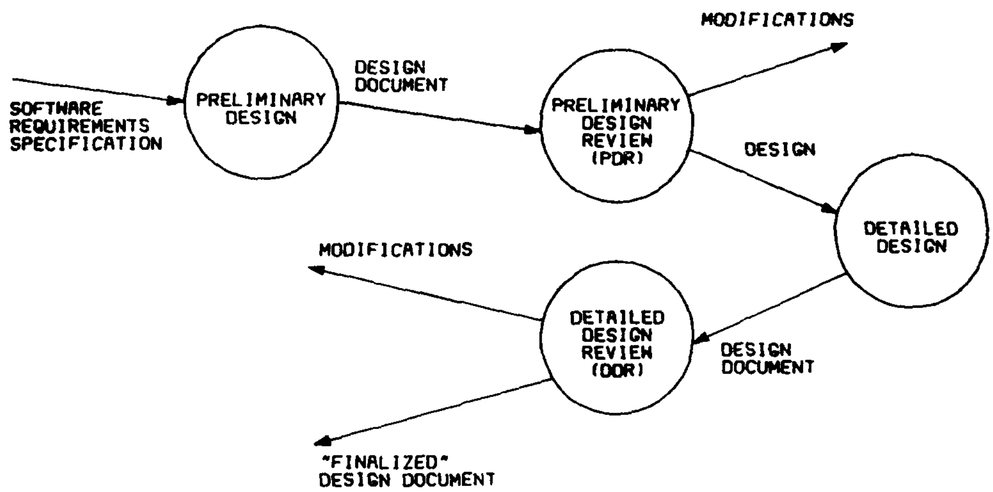


FIGURE 3. SOFTWARE DESIGN PROCESS

The propulsion control will include the closed loop control of the shaft RPM, propeller pitch and engine sequencing (e.g., start, stop, assume power).

Remote Terminal Unit Firmware

The Remote Terminal Unit provides a suite of software which allows the control and monitoring of the plant.

Control is always initiated when commanded by the DPC. Monitoring of the plant is carried out automatically and the resultant changes in the status of the plant transmitted to the DPC.

SYSTEM TEST

The system (both hardware and software) tests will be conducted in two steps.

- Unit tests - these tests focus on each module individually, assuring that it functions properly as a unit.
- Integration tests - these test focus on a complete system. These tests are used to provide assurance that the system meets all the functional and performance requirements.

A bottom-up process will be followed in the integration tests of both the software and hardware.

CONCLUSION

The IMCS which is based on the SHINMACS concept incorporates an advanced and highly effective man-machine interface. It stands at the forefront of warship propulsion control and provides the performance, reliability and survivability required for a warship machinery control system.

REFERENCES

(1) Human Engineering Design Requirements for SHINMACS Machinery Control Consoles, Part III Console Displays and Operation, DCIEM Technical Report 81-R-18, Toronto 1981.

(2) P. Penny, Lt-Cdr. D. Marshall, "A Shipboard Integrated Marine Control System" Canadian Shipping and Marine Engineering, Vol. 5, No. 6, 1984, pp 22-24.

(3) CAE Electronics Ltd., "Integrated Machine Control System for the Canadian Patrol Frigate", TPD 8784, Montreal 1984.

(4) Roger S. Pressman, "Software Engineering: A Practitioner's Approach." McGraw Hill Series in Software Engineering and Technology, New-York.

(5) UNIVAC Defence Systems, "Serial Data Bus Subsystem User's Manual for SHINPADS Engineering Development", St. Paul, Minn., USA, March 1983.

THE PRACTICAL CONSIDERATIONS OF DESIGNING A MICROPROCESSOR BASED CONTROL AND SURVEILLANCE SYSTEM

by Douglas A. Alsop
Teddington Industrial Equipment Ltd.

ABSTRACT

When considering any new design for a control and surveillance system the use of a microprocessor based unit is one of the most attractive solutions currently available. There are many benefits offered by adopting this technology, but also many pitfalls. Teddington Industrial Equipment have recently introduced a microprocessor based control and surveillance system and this paper sets out to review some of the decisions that must be made and specific problems to be solved when applying microprocessors in a naval machinery space environment.

INTRODUCTION

For many years the auxiliary machinery plant in ships has been controlled and protected by electro-mechanical and, more recently, electronic based systems. These units have been developed over the years to give safe and reliable operation to enable the plant being controlled to perform the basic function of providing it's service to the ship. These have generally been stand alone units with hardwired links to a centrallised control system to provide remote start/stop and plant status.

The current trend towards digital control throughout the ship, lower manning levels, and greater emphasis on machinery health monitoring has created a requirement for a new design of controller. These units must be self contained and situated close to the plant to perform the function of a local control panel and to be capable of starting, stopping, monitoring, controlling and protecting the plant in much the same manner as the conventional systems. They must interface with various types of analog and digital input transducers, provide plant status and alarm annunciation displays and to have an output capability to control actuators, interlocks, solenoids and motor starters. In addition, they must be able to interface with a digital main surveillance system or data communication network.

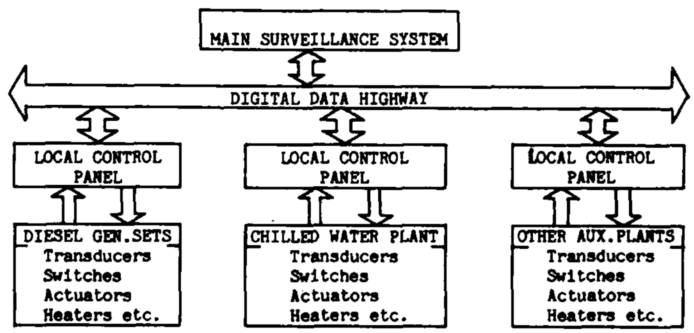


Figure 1. Local Control Panels & Data Highway

THE GENERAL REQUIREMENT

In formulating a specification for a general purpose controller suitable for naval and commercial use, there are many basic aspects, both technical and financial, which initially seem to conflict. One major dilemma is that generally, if equipment is specifically designed to meet the higher specifications required for naval use, then it tends to be too expensive for normal commercial applications. Conversely, standard commercial equipment often needs considerable bolstering to enable it to meet naval standards.

The requirement for the controller to be general purpose brings it's own particular problems for the designer in that it is difficult to determine how much capability the final design should have. This is another crucial financial factor where designs with large capabilities can be too expensive for small applications and units with less capabilities may be inadequate for larger applications.

If a design was to be based on discreet analog and digital devices, then the flexibility of such a system would usually be dependent upon either a number of purpose-built hardware modules, or a smaller selection of versatile hardware modules that, when applied, would have a percentage of redundancy.

When considering a microprocessor based design the inherent flexibility of the software is an enormous asset. The hardware design is much simplified because it is reduced to packaging the necessary microprocessors, memory, peripherals and providing standard methods of interfacing with the operator's console and the inputs and outputs of the plant to be controlled. A microprocessor based design has a lower component count than a discreet system and this has obvious benefits in the reduction of overall size, weight and cost of the final unit. The requirement to be able to interface with other digital systems or data highways can be readily implemented and other very important and useful facilities, such as self check and test routines, can be incorporated with only the minimum of extra hardware.

A major problem when considering the design of a microprocessor based system is the high cost of initial development and the technical risk involved. These can be minimised by utilising existing packages of hardware and software, when available. Some programmable logic controllers are well suited to the requirement in that they can process analog and digital inputs and outputs and can be easily expanded to cater for larger systems. However, the majority of them are designed and packaged for normal industrial applications and would not be suitable for naval use, where high ambient temperature and a shock capability is required. In addition, some design and development would be needed to create an operator's console and if this had to be done using the readily available input/output blocks, then the final cost would be significantly higher than a purpose built unit.

There are various types and styles of single board computer systems which could be adopted. These offer the designer a range of various CPU's, memories and analog and digital input/output units. Unfortunately, in many cases it is not possible to get precisely what is required off the shelf from one manufacturer and, due to inconsistencies in backplane allocation, boards from different sources are rarely compatible. It is then necessary to develop special purpose boards and the end design would be a hybrid.

From the designer's point of view with both the commercial PLC and the single board computer there could be considerable risk involving the long term availability of the bought out sub-assemblies.

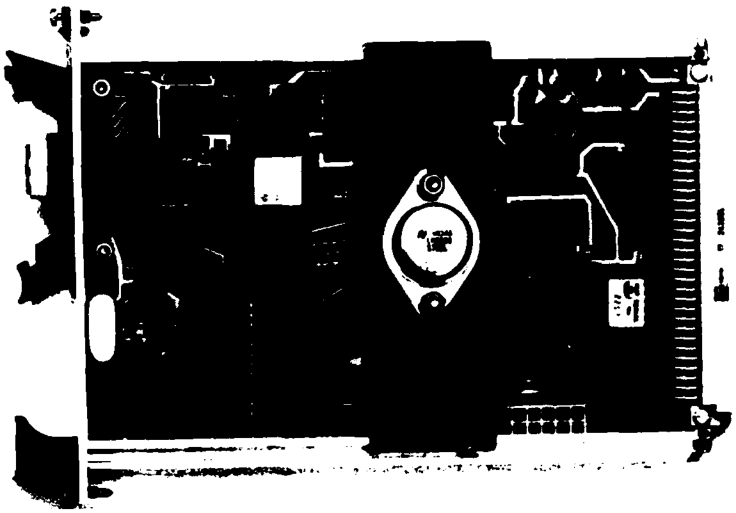


Figure 2. A Single Board Computer

THE INITIAL DESIGN STUDY

Teddington Industrial Equipment Ltd., having reviewed the available systems, decided that their best solution was to design and build their own system.

Based on the above-mentioned statements, the overall concept was to produce a microprocessor based controller that had expansion capabilities for input/output and a selection of basic tasks that could be called upon, as and when required. As the potential applications had a number of common functions, these could also be included in the standard features and tasks.

The operator's console could be standard for all applications, but being a costly piece of hardware to develop, careful attention was needed in the design stages to ensure it's flexibility.

The design criteria that was adopted with regard to the commercial/military options was:-

1. To select, wherever possible, devices that had pin compatable military and commercial variants.

2. Where military and commercial variants were not compatible to ensure that the PCB's could cater for either device by use of solder links, jumpering or other techniques.

3. If necessary, produce two types of PCB - one commercial and one military - that would perform identical functions.

4. Carry out full functional and environmental tests on any commercial equipment that might be considered.

The necessity to be able to easily reconfigure the system to suit various applications and, for possible field modifications, was to be catered for by the inclusion of an area of non-volatile RAM in the memory block. This reconfiguration capability was included in the software by the formation of a special applications program, which had the ability to allocate all external functions a dedicated internal reference upon which the standard tasks of the unit could perform actions.

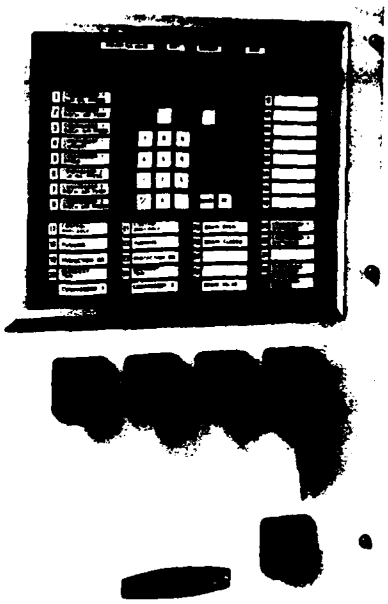


Figure 3. A Stand Alone Microprocessor Based Control& Surveillance System

To summarise, the aim was to produce a robust programmable control and surveillance system. It would be dedicated to it's role of control and surveillance and, therefore, be more cost effective but not as flexible as a programmable logic controller. It would however have more flexibility and be more cost effective than a discreet based unit.

HARDWARE DESIGN

Microprocessor Type

The first task to be tackled with any micro based hardware design is the selection of the type of processor. Currently available are 8-bit, 16-bit, and 32-bit devices. The choice depends greatly upon the speed of operation that is required and the amount of processing that must be carried out. Generally, the 8-bit devices, peripherals and memories are the least expensive but the slowest with the 16-bit devices being faster and more expensive etc.

The choice of microprocessor family is another difficult decision. To compare them in isolation by a bench marking technique is not altogether satisfactory ,ie the particular advantages of one type will not necessarily be the most benefical to the required application. The overall operation of the envisaged scheme, whether it will be required to operate by polling, interupts or a combination of both must also be considered, as this is an area where microprocessor families differ considerably. Other factors in the decision making are the long-term availability of the devices and what experience and development aids are readily accessible in house for a particular type. The capability of the relevant interface devices for a particular family are important, especially for an application with a considerable amount of input/output processing as their role will be very significant.

Memory Types

For a device to operate in the harsh conditions envisaged, it is of paramount importance that the integrity of the memory devices is of the highest standard available. With the ROM memory this is probably best achieved by adopting the mask programmable technique, where the devices are supplied direct from the manufacturer to the designer's specification. However, this solution is only really viable where quantities of 5,000 plus are involved. The unit cost of the devices is relatively low, but the masking charge and minimum order quantites are considerable.

An alternative solution is the fusible link PROM, where the memory devices are purchased blank and are subsequently blown to the designer's specification by passing a high current to rupture internal fusible links. These devices are more expensive per unit than mask programmed devices, but the programming of them is relatively simple and can be carried out either by a distributor or in house, with a suitable PROM programming device. For initial development work the ROM memory can be stored on U.V. erasible devices, but due to environmental limitations these would not be suitable for use on the final design.

The selection of the type of RAM memory is simpler as it only has to conform to the required speed and environmental specifications.

As previously stated, the reconfigurability of the system is an important factor and hence, memory must be provided to facilitate this. Some PLC's have their application program stored on tape and then download it into RAM on initial commissioning, but due to the RAM being volatile it is necessary to ensure that an emergency power supply or battery back-up is provided to prevent inadvertent loss of this data.

Another approach is to produce a dedicated fusible link PROM for each application to store this data. This solution however reduces the ability for the operator to make adjustments to threshold levels and process times. This requirement can be achieved to a limited extent by holding default values in the ROM, which are downloaded to RAM on initialisation. These could then be altered by the operator, but would return to the default state on any subsequent power up.

An ideal device for this requirement is the EEPROM (Electrically Erasible Programmable Read Only Memory). These have been available for some time, but recent developments have produced devices that require only a 5 volt power supply to operate, making them even more attractive to the designer. These devices can be externally interlocked to ensure that they can only be read from and not written to inadvertently. They are non volatile and hence need no battery back-up.

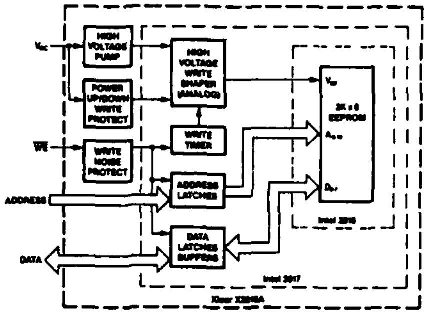


Figure 4. The Evolution of EEPROMS

Digital Input/Output

For any electronic equipment to be installed in the vicinity of industrial machinery, special attention is needed at the interface of the field transducers and actuators. This is best accomplished by adequate filtering (either hardware, software or both) and, where possible, complete isolation. Opto isolators are most commonly used for this purpose and there is a wide range of various types available to suit the speed and power requirements.

For outputs various options present themselves, such as transistor switching, triacs, solid state or electro mechanical relays. These would be selected depending upon the type of actuator with which they are to be interfaced.

Analog Input/Output

Analog inputs are prone to interference, pick-up and power supply transients etc. and attention must be given to the screening, filtering and isolation of the inputs. With a micro design various scaling and offsets can be incorporated in the software and these can be combined with a purpose built conditioning unit to enable various styles of transducer to be used in the final system.

Analog to digital conversion can be handled in various ways, depending upon the accuracy, resolution and speed required for the overall system. By converting the analog signal to a frequency the processor can handle this as digital information and perform the conversion itself. Alternatively, a proprietary discreet A/D converter could be employed to relieve the processor of some of it's duties. The requirement to handle a number of analog inputs can be provided by the use of multiplexors. If this process is adopted, then the polling rate of the analog inputs will become slower as more analog inputs are required. To redress this a solution is to offload the analog handling requirements to a dedicated sub-processor, relieving the main processor of a considerable amount of it's duties.

Analog outputs, if required, could be provided with the use of proprietary D/A convertors.

The Operator's Console

The operator's console has to be simple to operate, bearing in mind that it will be used by plant machinery maintainers, who may not necessarily be aware of computer operating techniques. It must be robust enough to withstand the machinery space environment, where it could come into contact with oil and firewater sprinklers and sprays. Some totally sealed keypads have high specifications, but tend to be expensive. Touch sensitive or membrane switches can be produced with an adequate specification, but for operational safety they would not be considered suitable for any function which may cause an executive action to occur.

There are a diversity of numeric and alpha numeric displays currently available and the selection of a particular type would depend on the requirement. Generally LED's are considered suitable for a dim ambient light situation, but are less discernible in broad daylight. Liquid crystals consume far less power than LED's, but would require back-lighting for low ambient light conditions. Both types of display are available to operate from a single power supply and have a high degree of reliability.

The final design adopted by Teddington Industrial Equipment for their product "Series 6" is briefly described below.

The concept was to have a central processing unit CPU, power supply monitor PSM, and operator's console as the basic unit. To these could be added hardware modules for analogue inputs, digital input/output, data transmission etc.

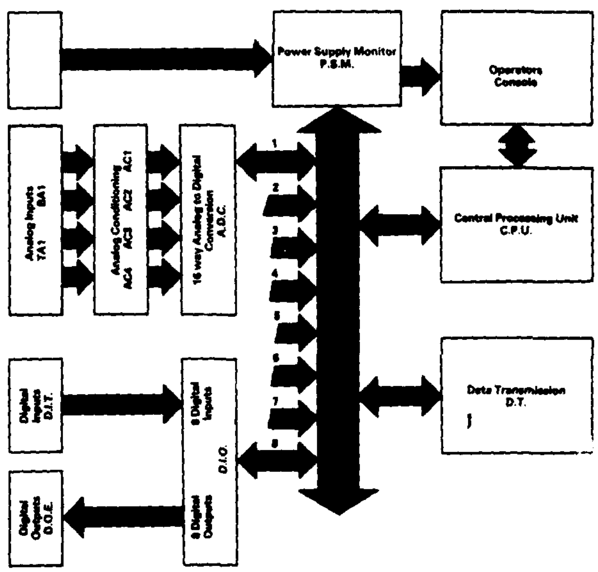


Figure 5. Series 6 Hardware Block Diagram

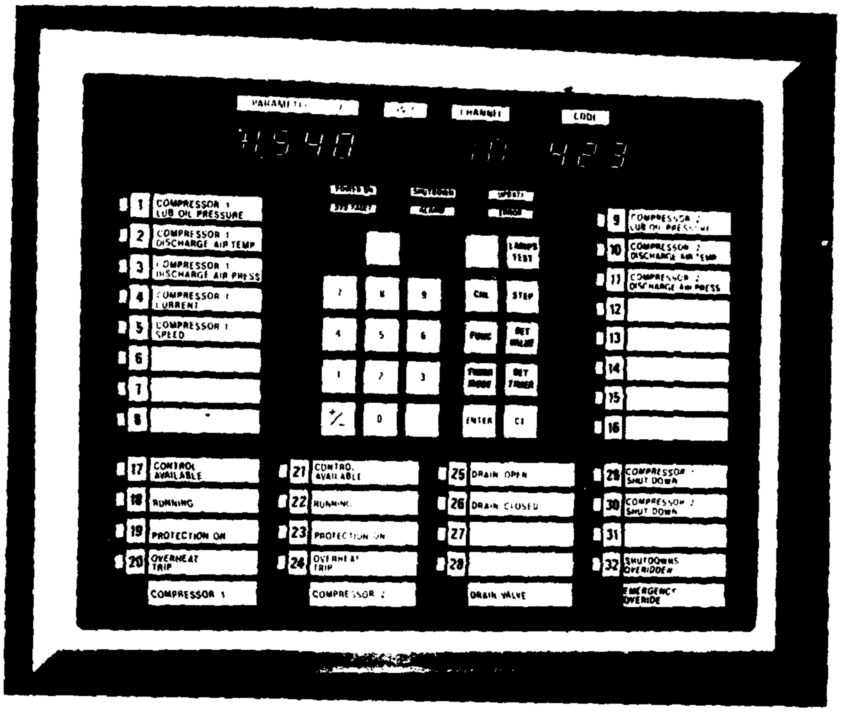


Figure 6. The Operator's Console

THE SOFTWARE

The software for the Teddington Series 6 was designed to allow a combination of 8 Input/Output boards, either digital or analog, to be operated at any one time. The actions taken by the controller are governed by an applications program, which is written specifically for each type of application and stored in the EEPROM. The applications program is interpreted by the operating system in the ROM and executed sequentially.

The configuration tables

All inputs and outputs connected to the system are allocated internal references. The software is so written that these internal references can refer to either analog or digital inputs or outputs. Their actual status is defined in tables in the EEPROM and their allocation is part of the application configuration data. Also stored in the configuration data tables are any scaling and offset conditioning factors required by analog inputs, and the initial default values (normally energised, normally de-energised) of the digital inputs and outputs.

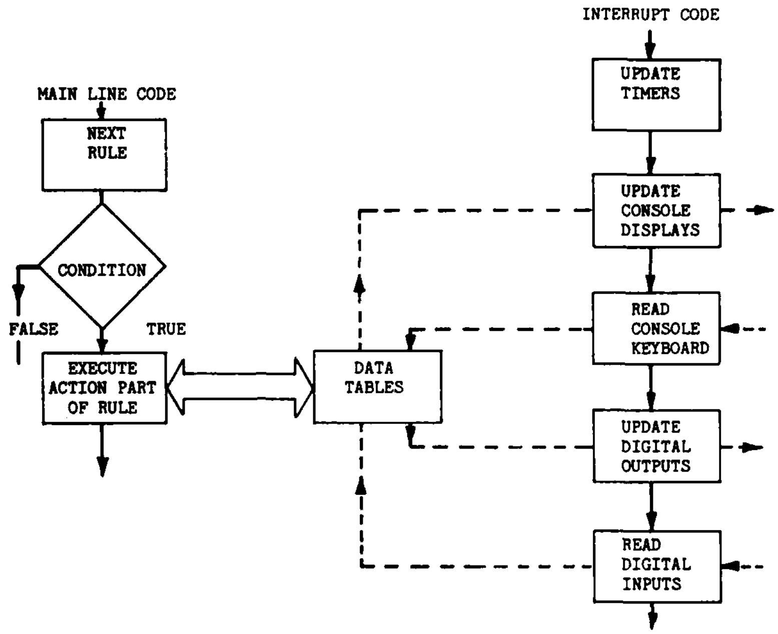


Figure 7. Part of the Main Line Rules & Interrupt Flow Chart

The operating system

The operating system utilises a 4 millisecond interrupt, which is generated from an internal timer. This interrupt controls the necessary clocks used for timing purposes and the rate of multiplexing on the operator's console displays and keypads.

In normal operation the software sequentially reads a block of applications program which consists of a number of rules. Any executive action required to be taken as determined by those rules is stored in the data tables. The 4 millisecond interrupts of the system enable the data tables to be read from to carry out external actions and to be written to for updating of input information.

The applications program

The application engineer's first task when applying the system is to allocate the internal references to the required inputs and outputs and enter these into the configuration data tables.

The applications program is then created by applying a rule to each action required. The rules typically enable outputs to be switched on and off, timers to be started and stopped, warning, shutdown or system fault annunciation sequences to be initiated etc. Each rule has the ability to consider the status of a subject, compare it with a pre-set value (if required), carry out various actions, change the status of a target and, if necessary, invoke a time delay.

These rules, once entered, are stored in the EEPROM and cannot be altered without a special programming device. However, the pre-set value and time delay parts of the rule, where applicable, can be altered from the operator's console under the control of the lockable key switch.

This method of programming enables any number of rules to be applied to any input, allowing multiple threshold levels to be set and multiple actions performed. A typical example of rule programming is shown in fig. 8

In this example it is required to monitor an analog input and initiate an alarm, if it deviates above the threshold level. However, a delay is required so that transient excursions of less than n seconds will be ignored. This requirement takes two rules.

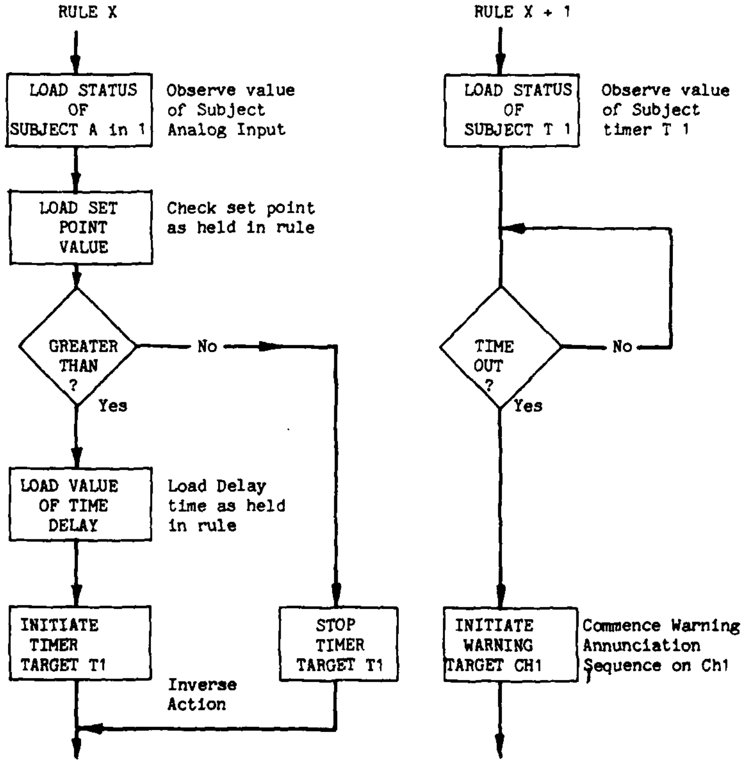


Figure 8. Typical Example of Rule Programming

The applications program is normally written on a micro computer with the aid of an assembler program and then downloaded into the EEPROM. It can also be loaded with a hand held programming device directly into the machine to facilitate field modifications or reconfiguration.

Special purpose programs

On initial system power up the controller continuously operates the mainline code and executes the actions as directed by the applications program. For test and system diagnosis however, the controller can be driven into special test routines with the aid of the hand held programmer. These can check the functions of the CPU, memory, console and all input/outputs. In addition, the machine can be operated so that all addresses and data are presented on the operator's console. This feature enables the service or commissioning engineer to make minor reconfiguration changes when on site.

CONCLUSIONS

Microprocessor based controllers require special attention to their design and development to enable them to be used in naval machinery space environments. Once this has been achieved then the inherent flexibility of the controller is a significant asset. When compared against discreet systems the micro design's lower cost, smaller size and weight are advantageous to modern ship design. Alternatively, those savings could be used by making the machine monitor more parameters, or carry out more sophisticated duties to assist in improved plant efficiency, trend analysis and machinery health monitoring.

The ability to communicate intelligently with a centrallised control room greatly assists in general watchkeeping duties, but above all, the unit's independence from the data link allows the greatest aspect of safety during damage control situations and facilitates local plant fault finding and commissioning.

DESIGN ASPECTS OF A SUBMERSIBLE STEERING CONTROL SYSTEM

by D. Beekman, G.B.H. Jacobs and B. Twigt
Applied Dynamics Europe
a department of
Van Rietschoten & Houwens E.M. B.V.
Rotterdam, The Netherlands

ABSTRACT

This paper describes the design of a steering system for control of the course and the depth of a submersible.

The control algorithms will be implemented into a general purpose military computer system which enables application of advanced control techniques like :

- optimal linear quadratic control
- feed forward control and mode switching
- gain scheduling for the speed
- adaptive Kalman filtering
- estimation of slowly varying disturbances

The decoupled depth and course control algorithms use simplified control models, respectively based upon a fourth order linear approach of the depth behaviour and an extended first order Nomoto model for the course behaviour.

1. INTRODUCTION

The submersible for which the control algorithms are developed can be controlled by means of two sets of hydroplanes for depth and pitch and one rudder for course. The hydrodynamic coefficients of the vessel are known, both via model tests and trials with similar submersibles.
Although the design of a steering control system means more than the development and test of control algorithms, this paper is limited to those items.
The work that V. Amerongen (Ref. 1) has done in the past and still is doing in this field has been of great importance to us. Use of modern simulation equipment proved to be very advantageous. Although various techniques are available to develop control algorithms, good and fast, simulation facilities are of enormous help.

2. MODELLING

In order to develop and test the control algorithms in a realistic environment our first goal was to simulate the complete non linear 6 DOF equations of motion of the submersible as accurate as possible. (Fig. 1).

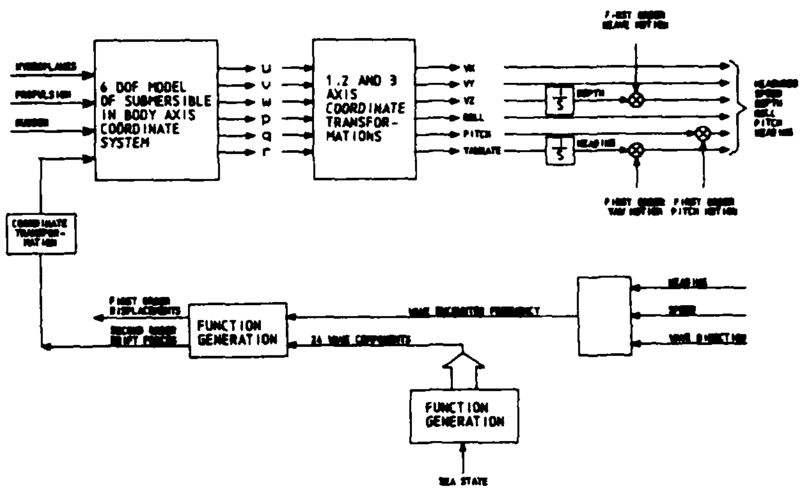


Fig. 1. Simulation set up of the submersible's equations of motion.

We used the equations of motion as given in Report 2510 of the Naval Ship Research and Development Center. From this model, we derived simplified models for depth and course which are independent of each other under the following assumptions :

- An approximately symmetrical submersible
- Small roll angles
- An approximately constant forward speed during course and depth changes

For the vertical plane, the equations of motion of the simplified model are :

$$\begin{bmatrix} \dot{z} \\ \ddot{z} \\ \dot{\theta} \\ \ddot{\theta} \end{bmatrix} = \begin{bmatrix} A \end{bmatrix} \begin{bmatrix} z \\ \dot{z} \\ \theta \\ \dot{\theta} \end{bmatrix} + \begin{bmatrix} B \end{bmatrix} \begin{bmatrix} db \\ ds \end{bmatrix} + \begin{bmatrix} C \end{bmatrix} \cdot \begin{bmatrix} F \\ M \end{bmatrix}$$

Where the terms of the A, B and C matrices are a function of the hydrodynamic coefficients and (some) of the speed of the submersible. The outputs of this simplified model are close to those of the complete 6 DOF model even when large hydroplane angles are commanded. This model (fig. 2) is used for the development of the control algorithms for the depth controller.

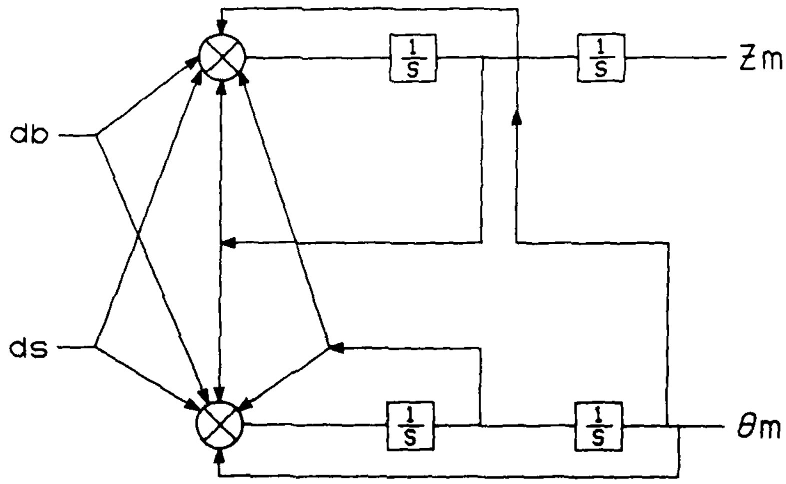


Fig. 2. Simplified model for the vertical plane

For the horizontal plane, a model was chosen which is valid for most rudder angles and the whole speed range.
The lateral speed, v, approximates a constant times the rate of change of the course. This can be computed from the equations for the lateral displacement and the yawing moment.

The equation for the lateral displacement under the assumptions that the angular velocity components relative to fluid about X and Y axis, p and q, are zero and that all accelerations are zero, simplifies to:

$$h_1 * u * r + h_2 * r * |v| + h_3 * u * v + h_4 * dv * u * u + h_5 * v * |v| = 0$$

substitution of $v = C * \dot{\psi}$, $r = \dot{\psi}$ and the values for the hydrodynamic coefficients results in:

$$-dv = \frac{1}{u} * \dot{\psi} + 60 * \frac{1}{u} * \frac{1}{u} * \dot{\psi} * |\dot{\psi}|$$

substitution of $r' = \frac{1}{u} * L * \dot{\psi}$ results in:

$$-dv = \frac{1}{L} * r' + \frac{60}{L * L} * r' * |r'|$$

This expression gives a good description of the relationship between the rudder angle dV and the variable r during course changing operations.
The equation for the yawing moment, under the assumption that $v = \dot{v} = 0$, is :

$$Iz * \dot{r} = h_8 * \dot{r} + h_7 * r + h_8 * dv$$

This is the well known first order Nomoto model, which is only valid, because of the assumption that was made, for small rudder angles.
Combining this Nomoto model with the simplified equation for the lateral displacement results in a model which is valid for all rudder angles and the whole speed range (Fig. 3).

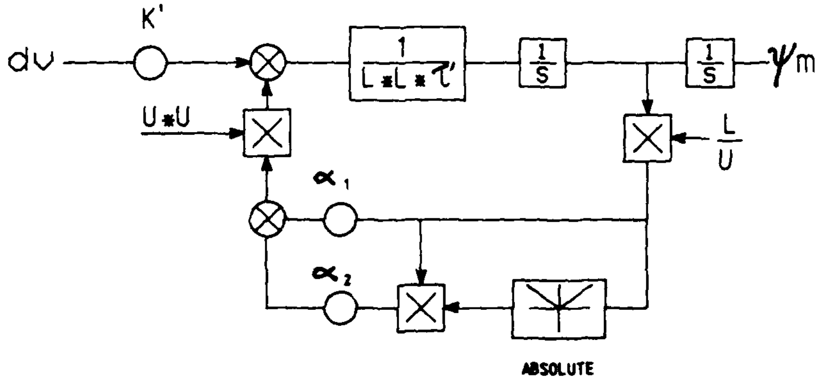


Fig. 3. Simplified model for the horizontal plane

3. MODELLING THE DISTURBANCES

In order to study the effect of waves on the submersible, we used the following approach :
the Pierson Moskowitz spectrum, defined as,

$$S(\omega) = \frac{A}{\omega^5} e^{-\frac{B}{\omega^4}}$$

$$A = 0.77915, \quad B = 3.11/Hs \cdot Hs$$

is divided into 24 bands, each with the same frequency interval $d\omega$.
Each band can be associated with a wave having the same energy and frequency as the band.
The wave elevation can then be described by a combination of the 24 waves using random phase angles ei.
The amplitude of a wave with frequency ωi is:

$$Ai = \sqrt{2S(\omega i)d\omega}$$

The wave elevation at the centre of gravity of the vessel is given by :

$$Z(t) = \sum_{i=1}^{24} Ai\cos(\omega i\, t(1 - \omega i\, v/g) + ei)$$

Where $\omega i(1-\omega i\, v/g)$ is the encounter frequency of the waves.
This value Z(t) is added as noise to the depth signal.
Other disturbing factors caused by waves are the first order wave forces and moments and the second order drift forces and moments.
The first order forces and moments are replaced by motions. Each wave component generates a motion in heave, pitch and yaw.
The amplitude and phase of the motion are given in tables.
By means of function generation with the wave direction and the speed of the vessel as inputs, the first order displacements are computed.
The instantaneous wave amplitude squared, A * A, and the instantaneous wave frequency ω are also computed. The drift forces and moments are given by $Fi(\omega)$ * A * A, where Fi is a table with either heave forces, pitch moments or yaw moments.
Since those forces and moments are given in an earth-fixed coordinate system, they have to be transformed to the body axis system and then used in the non linear 6 DOF equations of motions of the submersible (see fig. 1).

4. CONTROL IN THE VERTICAL PLANE

The submersible control takes place on several levels.

The first level consists of a linear feedback of the actual states, based on optimal control theory.
This feedback control reduces the state errors by ordering the proper plane deflections as a function of the error between the reference state and the actual state.
The feedback gains are computed in such a way that a quadratic costfunction is minimised. This cost function is chosen according to the specifications of the autopilot and the obtained simulation results.
Before the ordered plane deflections from the optimal controller are actually used, they are checked against the possibilities of the steering machine. If the ordered plane speeds exceed the speed limits of the steering machines the plane orders are adjusted in such a way that the commands are in agreement with the speeds of the steering machines. In order to compute the eight feedback gains (four for each plane since the state error vector consists of depth and pitch error and depth speed and pitch speed error) it is necessary to determine a linear model in the vertical plane of the submersible. Some of the terms of the A, B and C matrices of the model mentioned in the previous chapter depend upon the speed of the submersible.
In order to get matrices with constant coefficients, needed to compute the feedback gains, the speed range of the submersible is divided into a number of intervals and for each speed interval the eight feedback gains are computed by linear interpolation.

The second level of control is a feed forward controller. For depth changes it is possible to compute the necessary plane angles in order to achieve the desired values of pitch angle and the depth rate of change. The assumption that $p=q=r=\dot{p}=\dot{q}=\dot{r}=\dot{u}=\dot{v}=\dot{w}=v=0$, and that $w=0$, or $\dot{z}=u*\theta$ simplifies the non-linear equations for the normal force and the pitching moment into:

$$z_1*ds+z_2*db=0$$

$$m_1*\theta+m_2*ds+m_3*db=0$$

From these two equations the values for the plane angles, db and ds, can be computed during major depth changes where $w = 0$.

The feed forward control is also used in the case that a pitch angle is desired during depth keeping.
Again it is assumed that $p=q=r=\dot{p}=\dot{q}=\dot{r}=\dot{u}=\dot{v}=\dot{w}=v=0$
but in this case $w\neq0$.
The same, before mentioned, non linear equations now become :

$$z_1*ds*u*u+z_2*db*u*u+z_3*u*w+z_4*w*|w|=0$$

$$m_1*\theta*u*u+m_2*ds*u*u+m_3*db*u*u+m_4*u*w=0$$

Substituting $w = u * \theta$ and dividing by $u * u$ results in :

$$z_1 * ds + z_2 * db + z_3 * \theta + z_4 * \theta * |\theta| = 0$$

$$m_1 * \theta + m_2 * ds + m_3 * db + m_4 * \theta = 0$$

From these two equations the values for the hydroplane angles can be computed in order to maintain a certain pitch during depth keeping.

The third level of control consists of the computation of the (slowly) varying disturbing forces and moments and the necessary plane angles to counter act those forces and moments. This is done by comparing the state variables of the submersible with the states of a model of the submersible. The linear equations for the vertical plane are:

$$\ddot{z} = a_{22} * \dot{z} + a_{23} * \theta + a_{24} * \dot{\theta} + b_{21} * ds + b_{22} * db + c_{21} * F$$

$$\ddot{\theta} = a_{42} * \dot{z} + a_{43} * \theta + a_{44} * \dot{\theta} + b_{41} * ds + b_{42} * db + c_{41} * F + c_{42} * M$$

The equations for the model do not contain the F and M, the plane angles for the model are the measured plane angles from the submersible minus the correction angles.

$$\ddot{z}_m = a_{22} * \dot{z}_m + a_{23} * \theta_m + a_{24} * \dot{\theta}_m + b_{21} * (ds - \Delta ds) + b_{22} * (db - \Delta db)$$

$$\ddot{\theta}_m = a_{42} * \dot{z}_m + a_{43} * \theta_m + a_{44} * \dot{\theta}_m + b_{41} * (ds - \Delta ds) + b_{42} * (db - \Delta db)$$

substitute $\Delta ds = \Delta sds + \Delta\Delta ds$, $\Delta db = \Delta sdb + \Delta\Delta db$

In the 'pseudo stationary' case where all accelerations are Ø, the equations for the submersible are simplified to :

$$c_{21} * F = -b_{21} * \Delta sds - b_{22} * \Delta sdb$$

$$c_{41} * F + c_{42} * M = -b_{41} * \Delta sds - b_{42} * \Delta sdb$$

Subtracting the model equations from those of the submersible results in :

$$a_{22} * \Delta \dot{z} + a_{23} * \Delta \theta + a_{24} * \Delta \dot{\theta} + b_{21} * \Delta\Delta ds + b_{22} * \Delta\Delta db = 0$$

$$a_{42} * \Delta \dot{z} + a_{43} * \Delta \theta + a_{44} * \Delta \dot{\theta} + b_{41} * \Delta\Delta ds + b_{42} * \Delta\Delta db = 0$$

These equations can be solved for $\Delta\Delta$ db and $\Delta\Delta$ ds.
Integrating those values gives the offset plane angles Δds and Δdb. In order to make sure that the states of the model do not drift from the measured states of the submersible, parts of the error signals Δz and $\Delta \theta$ are used as inputs to the model to prevent this drifting.

5. CONTROL IN THE HORIZONTAL PLANE

The control of the submersible in the horizontal plane also takes place on three levels.

The first level consists of a PD controller.
For the proportional part of the controller, Kp is chosen as :

$$K_p = -2.5 * \frac{u_\emptyset}{u}, \quad -K_p \leqslant 5$$

where $u\emptyset$ is the cruising speed of the submersible. According to V. Amerongen (Ref. 1), a suitable Kd is the following :

$$K_d = \frac{L}{u} * (2 * d * \sqrt{K_p K' \tau' - 1}) / K'$$

where K' and τ' are the gain and time constant of the first order Nomoto model.
d, the damping, is chosen in such a way that good simulation results are achieved.

The second level also consists of a feed forward controller. The feed forward control consists of the function mentioned in the previous chapter :

$$-dv = \frac{1}{L} * r' + \frac{6\emptyset}{L * L} * r' * |r'|$$

The third level of control compares the course rate of change of the submersible with that of a model. The difference between the two values is used to compute the offset rudder and the disturbing moment.

6. THE CONTROLLER MODES

Both the depth and the course controller distinguish between two modes, depth (course) keeping and changing. In the keeping mode the setpoints for the depth controller consist of :

- desired depth
- desired rate of change of the depth, is ∅
- desired pitchangle during depth keeping
- desired rate of change of pitch, is ∅

For the course controller the setpoints are :

- desired heading
- desired rate of change of the course, is ∅

In the changing mode, the setpoints for the depth controller are :

- desired depth, is the measured depth
- desired rate of change of the depth, is u * diving angle
- desired pitchangle, is diving angle
- desired rate of change of pitch, is 0

For the course controller the setpoints are :

- desired heading, is the measured heading
- desired rate of change of the heading

So in the changing modes the depth error and the heading error are 0.

The point where the mode changes from changing to keeping is determined as follows :
During the changing mode the plane/rudder angles are almost completely generated by the feed forward controller, during the keeping mode it is the optimal controller which generates the angles. For the course controller, the rudder angle in the changing mode is:

$$dv = \alpha_1 * \dot{\psi}g + \alpha_2 * \dot{\psi}g * |\dot{\psi}g|$$

When the mode is switched to keeping, the PD controller generates in the first instant

$$dv' = K_p * \Delta\psi + K_d * \dot{\psi}g$$

with $\dot{\psi}g$ = desired rate of change of the heading
$\Delta\psi$ = heading error

The mode switching is optimal when dv = dv′, or

$$\Delta\psi = (\alpha_1 - K_d + \alpha_2 * |\dot{\psi}g|) * \dot{\psi}g / K_p$$

since $\alpha_1 \ll K_d$ and $\alpha_2 * |\dot{\psi}g| \ll K_d$, this results in:

$$\Delta\psi = -\frac{K_d}{K_p} * \dot{\psi}g$$

When during course changing the heading error becomes $\Delta\psi$, the setpoint vector is changed from the mode changing to the mode keeping.
The result is that no large, unnecessary jumps in the rudder angle occur.
For the depth controller, the optimal point to switch the mode is somewhat different for the bowplane and the sternplane, but averaging the two values gives satisfactory results.

7. FILTERING AND ESTIMATION OF STATE VARIABLES

Besides the slowly varying disturbances, which can be compensated for, there are also acting higher frequent disturbances on the submersible.
Noise due to waves, forces and moments proportional to the wave motions are examples of this kind of disturbances. The disturbances cannot be compensated for. The submersible acts as a low pass filter, it cannot follow the high frequent components of the rudder and plane motions.
In fact, high frequent rudder and plane motions do not have any positive effect on the reduction of setpoint errors. Therefore it is necessary to filter the measured signals in order to suppress the high frequent rudder and plane motions.
For this purpose the before mentioned model of the submersible is used in the control computer. This model is actuated by the actual measured plane (rudder) angles and the calculated slowly varying disturbances.
The states of this model and the measured states of the submersible are weighed -in a certain proportion- in order to obtain filtered system states.
This is done by determining the low frequency energy component of the difference of the model and measured states and the total energy of the difference. Dividing the low frequency energy by the total energy gives the weighing factor to obtain the filtered states.
The weighing factor is also used as feedback gain in the model, which has the effect that the model is now a noise adaptive Kalman filter.
The measured states consist of the depth, pitch and the course. For the optimal depth controller and the PD course controller it is necessary to obtain the rates of change of those states as well.
The noise adaptive Kalman filter provide these signals.
In fig. 4 is illustrated how the computed feedback gain is used in the model of the submersible and how it is used in order to get filtered estimates of the state. Part of the model for the vertical plane is illustrated. The feedback value is KSID * (z-zm). In order to get the filtered estimate for the depth, $\hat{z}$, the model depth, zm, is added to the feedback value. This results in :

$$\hat{z} = KSID \times (z - z_m) + z_m$$

For low values of KSID, or much high frequency noise, the estimate is mainly dependent upon the model depth zm. When no noise is present, KSID = 1, then the estimate will be equal to the measured depth.
In order to get an estimate of the depth rate of change, $\hat{\dot{z}}$, the same feedback value is added to $\dot{z}$m, the depth rate of change of the model.

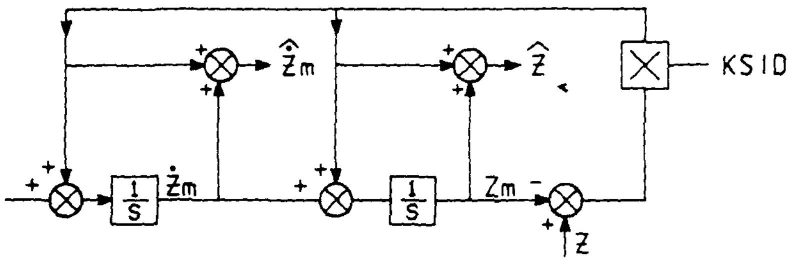


Fig. 4. Use of the weighing factor in order to get filtered states.

8. DETERMINATION OF THE FEEDBACK GAINS IN THE FILTER MODEL

The difference of the measured state variable and the corresponding model state variable is used in order to compute the low frequency component and the total energy in this error signal.
In order to compute the low frequency component the error signal is filtered in a low pass filter with a variable, helmsman definable, time constant.

The absolute value of this result is computed, then averaged over approx. 100 seconds and finally squared. To compute the total energy, the absolute value of the error signal is first averaged and then squared. Dividing the low frequency component by the total frequency component results in the feedback gain and the weighing factor of the filter model.

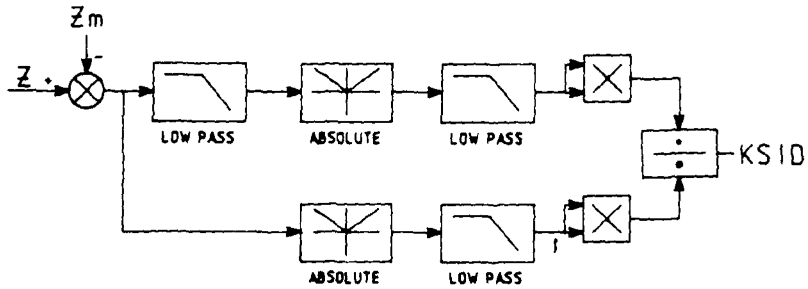


Fig. 5. Computation of feedback gain for the depth filter.

9. SIMULATION

In order to test the various control algorithms in a realistic environment, use has been made of various computer systems. As said before, our first goal was to simulate the complete 6 DOF equations of motion of the submersible as accurate as possible (ref. 3).
The next step was to incorporate the controllers in the simulation and the effects of waves and other disturbing factors as mis-trim and mis-ballast.
For detailed engineering purposes we used a PDP-11 system together with PSI, an interactive simulation package developed by Delft University of Technology.
This system has the capacity to simulate the 6 DOF model and various parts of the controllers faster than real time.
A second system which was used consists of a PDP-11 together with an AD-10 special purpose computer.
Due to the special architecture of the AD-10, a multi-processor system where each processor is dedicated to a specific task, it proved possible to simulate the 6 DOF model, all the controllers and the effect of waves approximately 200 times faster than real time. The simulation includes 164 functions of one variable, 165 functions of 2 variables, 280 summers, (each summer with up to 48 inputs), 4 coordinate transformations, 12 divisions, 80 integrators and 37 combined comparator/switches.
The frame time needed for one solution of all algebraic variables and the state variables is just over 2 milliseconds. The use of two independent systems proved to be highly advantageous. Apart from the high speed of the AD-10, comparing results of the two systems immediately points to the small programming errors which otherwise probably would have stayed unnoticed.

10. REFERENCES

(1) J. van Amerongen, "Adaptive Steering of Ships", Delft University of Technology, 1982

(2) *Naval Ship Research and Development Center*, "Report 2510"

(3) Applied Dynamics International, "Six-Degree-of-Freedom Flight Simulation Using the System 10", Application Report

SHIP STEERING CONTROL SYSTEMS MODELLING AND CONTROL DESIGN

by M.J. Grimble, R. Katebi, J. Wilkie
Industrial Control Unit
University of Strathclyde

ABSTRACT

A dynamic cost function is proposed for minimisation of propulsion losses due to steering. The cost function is based on the separation of the low and high frequency motions of a ship. The low frequency model consists of the ship model and the low frequency effect of the wave. The proposed cost criteria includes the coupling between yaw rate and sway velocity while being independent of the closed loop frequency of oscillation as is needed in most commonly used criteria.

The cost function is present both in time and frequency domains and the possibility of identifying the parameters of the cost criteria on line is discussed.

INTRODUCTION

Current interest in the design of adaptive autopilots has been stimulated by developments in modern optimal control theory, availability of cheap, computing power, the continuing rise in the price of fuel and the growth in the size of vessels, together with a greater acceptance of the use of sophisticated control techniques. The new design strategies include optimsation of simple PID algorithms [1], self-tuning regulators [2] adaptive filtering schemes with state feedback [3] and model reference adaptive regulators [4]. Most of the aforementioned design techniques require specification of the mathematical models of the ship and the disturbance forces, and a cost criteria which is to be optimised.

While the behaviour of the ship in calm water is relatively well established, the effect of sea disturbances on the control system has provoked much discussion and research in recent years [5]. The research mainly centres on the possible methods of implementing the wave model in the control design.

The effect of the waves on the ship can be considered to be of relatively high frequency, though in some cases, for example, following seas, the wave spectrum overlaps substantially the low frequency rudder control system bandwidth.

The wave disturbance can be included in the design using either of two current approaches. That is, the wave action is modelled by an input disturbance [6] or alternatively by an output disturbance [7] to the ship model. Which model structure is appropriate is determined by the control requirements; for example, the need to minimise only low frequency heading errors. A comparison between the two approaches is made in Section 1.

One of the main problems in autopilot design is the choice of the cost function. As mentioned by Clarke [8] the absence of sway velocity from the performance criteria used in recent autopilot designs, raises doubts as to whether the new autopilots actually minimise the consumption of fuel.

A dynamic cost function is proposed here that takes into account the coupling and the phase difference between sway velocity and yaw rate. Furthermore, there is no need to know the closed loop frequency of oscillation as in the commonly used quadratic cost criteria. The cost function can be used both in rough and calm weather by changing the parameters of a transfer function.

The formulation of the cost criteria in the frequency domain is given in Section 3. A time-domain equivalent cost criteria is given in Section 4. The effect of the Wave on propulsion losses is discussed in Section 5.

THE SHIP AND DISTURBANCE MODELS

The wave model can be considered either as an input or an output disturbance of the low frequency model of the ship. The two different methods are shown in Fig.1 and 2. Each block represents a part of the ship model as follows:

A: Model of the steering machine:

$$\dot{\delta} = A_\delta \delta + B_\delta \delta_c \tag{1}$$

where δ_c is the rudder command and δ the rudder angle.

B: High frequency disturbance due to high frequency components of waves:

$$\dot{x}_h(t) = A_h x_h(t) + B_h \zeta_h \tag{2}$$

where ζ_h is white noise.

C: Low frequency disturbance due to wind and low frequency components of waves:

$$\dot{x}_d(t) = A_d x_d(t) + B_d \zeta_d$$

where ζ_d is white noise.

D: Low frequency model of the ship:

$$\dot{x}_\ell(t) = A_\ell x_\ell(t) + B_\ell \delta \tag{3}$$

and the output and measurement models by

$$y(t) = Cx(t)$$
$$z(t) = y(t) + v_o(t) \tag{4}$$

where x_ℓ represents the low frequency states: sway velocity, yaw rate, heading angle and rudder state, $[v, r, \psi, \delta]^T$. The measurement noise is denoted by v_o.

The state space representation of the models in Figs. 1 and 2 become:

System 1

$$\dot{x}(t) = A_s x(t) + B_s u + D\omega$$
$$y(t) = C_s x(t)$$
$$z(t) = y(t) + v_o(t)$$

where

$$A_s = \begin{bmatrix} A_\ell & 0 & C_h \\ 0 & A_d & 0 \\ 0 & 0 & A_h \end{bmatrix} \quad B_s = \begin{bmatrix} B_\ell \\ 0 \\ 0 \end{bmatrix} \quad D = \begin{bmatrix} 0 & 0 \\ B_d & 0 \\ 0 & B_h \end{bmatrix} \quad C_s = [C_\ell 0 0]$$

$$X = [x_\ell, x_d, x_h]^T, \quad u = \delta_c, \quad \omega = [\zeta_d, \zeta_h]^T$$

and the transfer functions follow as:

$$x_\ell(s) = (sI - A_\ell)^{-1}B_\ell\delta + (sI - A_\ell^{-1})C_h(sI - A_h)^{-1}B_h\zeta_h$$
$$x_d(s) = (sI - A_d)^{-1}B_d\zeta_d$$
$$x_h(s) = (sI - A_h)^{-1}B_h\zeta_h$$
$$z(s) = C_\ell x_\ell(s) + v_o(s)$$

<u>System 2</u>

$$\dot{x}(t) = A_s x(t) + B_s u + D\omega$$
$$y(t) = C_s x(t)$$
$$z(t) = y(t) + v_o(t)$$

where

$$A_s = \begin{bmatrix} A_\ell & 0 & 0 \\ 0 & A_d & 0 \\ 0 & 0 & A_h \end{bmatrix} \quad B_s = \begin{bmatrix} B_\ell \\ 0 \\ 0 \end{bmatrix} \quad D = \begin{bmatrix} 0 & 0 \\ B_d & 0 \\ 0 & B_h \end{bmatrix} \quad C_s = [C_\ell 0 C_h]$$

and the transfer functions

$$x_\ell(s) = (sI - A_\ell)^{-1}B_\ell\delta$$
$$x_d(s) = (sI - A_d)^{-1}B_d\zeta_d$$
$$x_h(s) = (sI - A_h)^{-1}B_h\zeta_h$$

and

$$z(s) = C_\ell x_\ell(s) + C_h x_h(s) + v_o(s)$$

The low frequency disturbance model representing wind, current and the drift component of the wave forces are also included, as shown in Figs. 1 and 2.

Note that the output y in Fig.1 contains the high frequency wave motion information. Some of the high frequency motion is reduced since the ship behaves like a low pass filter. It is usually thought undesirable to feedback signals with high frequency components. The rudder can not react to high frequencies since the effective range of rudder activity is about 0 to ·025 Hz. High frequency rudder variations also increase the rudder drag force. Modelling the ship motion in waves as shown in Fig.1 implies some feedback of high frequency motion. However, this approach is followed and justified by Reid [6] based upon the overriding importance of minimising his cost criteria (to minimise the added resistance of yawing and swaying due to waves). This approach results in some additional complexities in the control design.

The alternative method, as shown in Fig.2, is to separate the high and low frequency motions and to consider the high frequency motion as an output disturbance. However, studies carried out by Reid into the effect of sea disturbances on ship steering shows that for a high speed container ship, substantial wave energy is present over a range of low frequencies, corresponding to different encounter angles for seas aft of the beam. Hence, this separation may not be appropriate but is a subject for debate.

The modelling approach shown in Fig.2 has some advantages over that of Fig.1:

(1) The steady state Riccati equation, required to find the optimal gain vector, can be decomposed into low and high frequency parts. This reduces the amount of computation [9]. (2) If an adaptive filtering scheme is to be used, it is easier to separate the high and low frequency estimation functions [10].

The cost criteria is discussed next, based on the ship model represented in Fig.2.

COST FUNCTION

Central to the cost criteria proposed in the literature for autopilot design is the concept of the "added resistance of the ship due to steering". This relationship can be written as:

$$\Delta X = (m + X_{vr})vr + \frac{1}{2} X_{vv}v^2 + \frac{1}{2} X_{\delta\delta}\delta^2 \qquad (13)$$

where m is the mass of the ship, X_{vr}, X_{vv}, $X_{\delta\delta}$ are hydrodynamic derivatives, v, r, δ are sway velocity, yaw rate and rudder angle.

The mean of ΔX is given by:

$$\Delta\bar{X} = \lim_{T\to\infty} \frac{1}{2T} \int_o^T [(m + X_{vr}vr + \frac{1}{2} X_{vv}v^2 + \frac{1}{2} X_{\delta\delta}\delta^2] \, dt \qquad (14)$$

Normalising $(-\Delta\bar{X})$ yields:

$$J = -\Delta\bar{X}_{norm} = \lim_{T\to\infty} \frac{1}{2T} \int_o^T [-\lambda vr + \gamma v^2 + \delta^2] \, dt \qquad (15)$$

where

$$\lambda = \frac{2(m + X_{vr})}{X_{\delta\delta}} \quad \text{and} \quad \gamma = \frac{X_{vv}}{X_{\delta\delta}}$$

Yawing and swaying of the vessel is assumed to stem from either self-oscillations due to the steering system or the forced oscillation due to waves. Although these oscillations do not exactly follow a sinusoidal pattern, they may be approximated as regular yawing of a simple periodic form [8]. The yaw rate, sway velocity and rudder angle can then be represented as:

$$r = r_a \sin(\omega t + \phi_r)$$
$$v = v_a \sin(\omega t + \phi_v)$$
$$\delta = \delta_a \sin(\omega t + \phi_\delta) \qquad (16)$$

The term 'γv^2' in equation (15) is usually neglected [5] in comparison with the cross-coupled term and the rudder term. Hence by substituting equations (16) into cost criteria (15) the average added resistance can be written as:

$$J = -\lambda \left(\frac{v_a r_a}{2}\right) \cos(\phi_v - \phi_r) + \frac{1}{2} \delta_a^2 \qquad (17)$$

where λ is positive. Clearly, the resistance due to the coupling between sway and yaw depends n the phase difference $(\phi_v - \phi_r)$. For $0 < \phi_v - \phi_r < 3\pi/2$ or $3\pi/2 < \phi_v - \phi_r < 2\pi$, the coupling force is negative and for $\pi/2 < \phi_v - \phi_r < 3\pi/2$, the force is positive. It is, in general, very difficult to measure or compute the phase difference. Hence, the two extreme cases, i.e. $\cos(\phi_v - \phi_r) = 1$ or -1, are considered in the control design. The first case results in a thrust force and normally occurs in following seas, while the second case usually occurs due to selfoscillation in the steering gear system. The effect of the wave disturbace on the added resistance is further discussed in Section 6.

The cost criteria in the form of equation (5) is not appropriate for course keeping autopilot design since it is not a function of heading error. Reid [13] suggests a term $\epsilon\psi^2$ be added to (15) resulting in

$$J = \lim_{T\to\infty} \frac{1}{2T} \int_o^T [-\lambda rv + \gamma v^2 + \epsilon\psi^2 + \delta^2] \, dt \qquad (18)$$

There is of course the problem of choosing an appropriate value for ϵ. Furthermore, minimisation of (18) does not necessarily lead to minimisation of (15).

Koyama [11], Norrbin [12] and Reid [5] approximate (18) by a quadratic criteria. In place of the approximations to periodic yawing and swaying used in equation (16), they assume

$$v = kr \quad (19)$$

where k is a constant value and

$$r = \psi_a \omega \cos t(\omega t + \phi_\psi)$$

where ω is the closed-loop frequency of oscillation. This leads to:

$$J = \int_0^\infty (\alpha\psi^2 + \delta^2)dt, \quad \alpha = \lambda k \omega^2 \quad (20)$$

for the case where sway velocity and yaw rate are in antiphase. The values suggested for α in the literature vary immensely.

A summary of different types of cost function used in autopilot design is given in Table 1. Unfortunately, it is not possible to compute the different performance criteria proposed for a single vessel due to a lack of information. The variation in λ can, however, be seen from Table 1. This variation stems from the fact that k in equation (19) is a function of frequency and can not be approximated by a constant value. This is also pointed out by Clarke [8]. Furthermore, the value of α in (20) is a function of the closed-loop frequency of oscillation and this is usually not available.

A dynamic cost function is formulated in the next section which is a better representation of propulsion losses and is also suitable for control design.

	Cost Function	Value of λ	Type of Ship
Koyama[11]	$J = \int_0^\infty (\epsilon^2+\lambda\delta^2)dt$	$\lambda=10$	cargo vessel
Norrbin[12]	$J = \int_0^\infty (\psi^2+\lambda\delta^2)dt$	$\lambda=0.1$	cargo vessel
van Amerongen[4]	$J = \int_0^\infty (\epsilon^2+\lambda_1 r^2+\lambda_2\delta^2)dt$	$\lambda_1=15$, $\lambda_2=8$ $\lambda_1=1.8, \lambda_2=6$	tanker cargo vessel
Reid[5]	$J = \int_0^\infty (\lambda\psi^2+\delta^2)dt$	$\lambda=3251$	container
Reid[5]	$J = \int_0^\infty (-\lambda vr+\delta^2)dt$	$\lambda=34.19$	container
Blanke[13]	$J = \int_0^\infty (\lambda r^2+\beta\delta^2)dt$	$\lambda=39.8(u\ /u)^2$ $\beta=8.1\times10^{4}$	tanker
Kalström[14]	$J = \frac{1}{N}\sum_{t=1}^{N} [\psi(t+k+1 \mid t)-\psi_r]^2$	$\psi(\cdot)$ is predicted value of heading error	
Lim et al[15]	$J = \int (\epsilon^2+\lambda_1 r^2+\lambda\delta^2)dt$	$\lambda_1=20, \lambda_2=10$	marine class
Clarke[16]	$J = \int_0^\infty (A\psi^2+Br^2+c\delta^2)dt$	A=[illegible].5, B=78730 C=4.6	200k tanker (const.speed)

Table 1: A summary of different cost functions suggested for autopilot design

DYNAMIC COST FUNCTION

Assuming that the ship's low frequency motions are caused by rudder activity, the rate of turn r and the sway velocity v are closely related through [13],

$$\frac{v(s)}{r(s)} = G(s) = -k\,\frac{s+\tau_v}{s+\tau_r} \tag{21}$$

and the rate of turn and the heading through

$$\frac{r(s)}{\psi(s)} = s \tag{22}$$

The parameters of G(s) can be identified using the method described in ref [13]. The values given for τ_v and τ_r for a container ship are 0.14 and 0.039 and for a VLCC tanker 0.12 and 0.023. The basic assumption in deriving the cost function (20) is that for low frequencies G(s) can be approximated by:

$$G(s) = \frac{-k\tau_v}{s+\tau_r} \tag{24}$$

The approximation follows by assuming that the value of τ_r is very close to the range of rudder frequencies.

To include this effect in the cost function, the relation (14) can be transformed into frequency domain as follows:

$$J_{norm} = -\lambda'\,\overline{vr} + \gamma'\,\overline{v^2} + \overline{\delta^2} \tag{25}$$

using the relation (22) and (23) J_{norm} can be written in terms of spectral densities using the relations:

$$\overline{vr} = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty}(\Phi_{vr} + \Phi_{rv})/2 \;\; ds \tag{26}$$

$$\overline{v^2} = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty}\Phi_{vv}\; ds \tag{27}$$

$$\overline{\delta^2} = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty}\Phi_{\delta\delta}\; ds \tag{28}$$

where

$$\Phi_{vr} = G(s)\Phi_{rr}(s)\;, \quad \Phi_{rv} = \Phi_{rr}(s)G(-s) \tag{29}$$

$$\Phi_{vv} = G(s)G(-s)\Phi_{rr} \tag{30}$$

$$\Phi_{rr} = -s^2\Phi_{\psi\psi}(s) \tag{31}$$

and

$$J_{norm} = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty}\{[-\lambda'(G(s)+G(-s))/2 \;+\; \gamma'\;G(s)G(-s)](-s^2)\Phi_{\psi\psi} + \Phi_{\delta\delta}\}\;ds \tag{32}$$

The term multiplying $\Phi_{\psi\psi}$ can be simplified as:

$$\{\lambda'(G(s)+G(-s))/2 \;+\; \gamma' G(s)G(-s)\}(-s^2) = \frac{-\{\lambda' k\tau_v\tau_r + \gamma' k^2\tau_v^2\}s^2}{(s+\tau_r)(-s+\tau_r)}$$

This, in turn, can be expressed as

$$H(s)H(-s) = \frac{-c^2 s^2}{(s+\tau_r)(-s+\tau_r)}$$

where

$$H(s) = \frac{cs}{s+\tau_r}$$

and $c \triangleq (\lambda' k\tau_V\tau_r + \gamma' k^2\tau_V^2)^{\frac{1}{2}}$

The cost function can now be written as:

$$J = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty}\left[H(s)H(-s)\Phi_{\psi\psi} + \Phi_{\delta\delta}\right] ds \qquad (33)$$

The advantage of using (33) is that there is no need to know the closed loop frequency of oscillation as in (20). Furthermore, the phase difference between sway velocity and yaw rate is automatically taken into account through H(s).

The disadvantage of (33) is that although it is based on minimisation of control energy, there is no weighting on heading angle and hence on distance sailed. It is proposed, in considering the tracking requirements of a vessel, that a term dependent on ψ^2 be included to give direct control on heading.

TRANSFORMATION OF THE PERFORMANCE CRITERION INTO TIME DOMAIN

The cost function (33) can be transformed into the time domain using the augmented states concept. For the system shown in Fig.3 the new states can be introduced as follows:

Assume that W in Fig.3 represents Nomoto's model for the ship

$$\psi = \frac{r}{s} \qquad (34)$$

$$r = \frac{K_n}{(\tau s+1)}\delta \qquad (35)$$

$$\tilde{\psi} = \frac{cs}{(\tau_r s+1)}\psi \qquad (36)$$

Transform (34) to (36) into state space equations:

$$\dot{\psi} = r \qquad (37)$$

$$\dot{r} = (K_n\delta - r)/\tau \qquad (38)$$

$$\dot{\tilde{\psi}} = (cr - \tilde{\psi})/\tau_r \qquad (39)$$

and the performance criteria in the time domain is given as:

$$J = \int_0^\infty (\tilde{x}^T Q\tilde{x} + u^T Ru)\, dt \qquad (40)$$

where $u = \delta$, $R = 1$

$\tilde{x} = [\psi, r, \tilde{\psi}]$

and

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Note the augmented system can be written in the standard state space form:

$$\dot{x} = Ax + Bu \qquad (41)$$

It is interesting to interpret $\bar{\Psi}$ as a fictitious heading which includes the effect of cross coupling between sway and yaw and the variation with frequency. For $\tau_r = 0$, J in (40) reduces to the standard cost criteria used by many authors in recent years.

THE EFFECT OF WAVES ON PROPULSION LOSSES

Using the relations (16), the effect of waves on the added resistance can be computed for different frequencies and encounter angles. The effect of the seaway disturbance on the ship steering problem of a 250000 twd tanker is given by Reid [17]. The disturbance is approximated by integration of wave pressure on the local section along the longitudinal axis of ship's hull and the sway force and yaw rate is computed by numerical integration. This data has been used here to plot the variation of $\cos(\phi_v - \phi_r)$ and $H_\omega(s) = (v(s)/r(s))\cos(\phi_v - \phi_r)$ for different frequencies and encounter angles. The results are shown in Fig.4 and 5.

It may be seen from Fig.4 that the phase difference between sway velocity and yaw rate at low frequencies is nearly constant. For the same range of frequencies [0, .02], the magnitude of v/r from Fig.5 is also small (compared to high frequencies). It is common practice to avoid the violent rudder action by blocking the high frequency modes in the feedback path. Although some authors [6] believe that feeding back the high frequency motion can improve the steering losses, this is not as practical if considering the present mechanical limitations of the steering system of the ships.

The low frequency effect of waves on the ship is not neglected since this can be modelled as an input disturbance as described in Section 2. Hence the same cost function as in (33) or (40) can be used for control design in rough seas. Varying the value of c and τ_r in H(s) would allow for different sea conditions and for increased weighting on the high frequency motions in performance criteria (33). This can be achieved by fitting a 1st order transfer function to the graphs of v(s)/r(s) shown in Fig.5 for different encounter angles.

CONTROL DESIGN TECHNIQUE

The controller can be designed using either the time domain version of the cost function (40) or the frequency domain version (33). In the former case, the control problem is an LQG problem which relies on solving the standard Riccati equation. In the latter case, the control design techniques developed recently by Grimble [18] can be used. The control design is already under investigation and it will be reported in later reports.

The proposed cost function involves only measurement of the yaw rate and heading angle. Most of the ships with conventional autopilots have the facilities for this purpose. Hence, for a self-tuning scheme, the number of parameters to be identified is kept at a minimum. In addition, it may be possible to identify the parameter c which determines the propulsion losses on line. The necessity of identifying c is mentioned by Clarke [8]. The parameter c should be identifiable, from the yaw rate measurement. This however, requires further investigation.

CONCLUSIONS

The two approaches of interpreting the wave effect as an input or output disturbance have been described. Due to the high frequency nature of the wave effect it is usually not considered desirable to feed this to the rudder which can only respond to low frequencies. However, due to the presence of some wave energy at the low frequency end of the spectrum it is not realistic to consider the wave motions as only being composed of high frequencies. Therefore, a model (Fig.2) is proposed which decomposes the HF and LF energy of the waves and where the LF is to be considered in the control design.

The abundant existing cost function derive from variations on the theme of "added resistance due to steering". However, due to the different assumptions made in accommodatng for the sway velocity and phase difference between sway and yaw, there is wide dissension as to the value of λ and as to what is actually being costed.

After the various existing cost functions were analysed, a dynamic cost function was proposed. Its advantage over the time domain counterparts is that there is no need to know the closed loop frequency of oscillation. Furthermore, the effect of coupling between yaw and sway is taken into account. It may be possible to identify the parameters of the cost function on line.

Finally, the effect of the wave is also considered and shown that the bulk of change in the yaw rate and sway velocity is in the high frequency region and can not be counteracted by rudder action.

REFERENCES

(1) M.I. Bech, "Some aspects of the stability of automatic course control of ships", J. of Mech. Eng. Sc., Vol. 14, No. 7, 1972.
(2) C.G. Kallström, K.J. Åström, K.J., N.E. Thorell, J. Eriksen, L. Sten, "Adaptive autopilots for tankers", Automatica, Vol. 15, No. 3, 1979.
(3) S. Saelid, N.A. Jenssen, J.G. Balchen, "Design and analysis of a dynamic positionng system based on Kalman filtering and optimal ntrol", Trans. IEEE AC, Vol AC-28, No. 3, March 1983.
(4) J. van Amerongen, "Adaptive steering of ships", PhD thesis, Delft University of Technology, 1982.
(5) R.E. Reid, "A proposal for performance criteria for propulsion losses due to ship steering". J. of Dynamic Systems, Measurement & Control, Vol. 104, June 1982.
(6) A.K. Tugcu, O. Coskunoglu and R.E. Reid, "Performance criteria sensitivity analysis of ship steering models with respect to shaping filter design parameters", ACC, 1984.
(7) S. Saelid and N.A. Jenssen, "Adaptive autopilot with wave filter", 6th IFAC Symp., Arlington, Virginia, USA, June 1982.
(8) D. Clarke, "Do autopilots swave fueld", Trans IMarE(c), Vol. 94, 1982.
(9) M.J. Grimble, "Design of optimal stochastic regulating systems including integral actin", Proc IEE, Vol. 126, No. 9, Sept. 1979.
(10) P. Fung and M.J. Grimble, "Dynamic positioning using a self-tuning Kalman filter", Trans IEEE AC, Vol. AC-28, No. 3, March 1983.
(11) T. Koyama, "On the optimum automatic steering system of ships at sea", J. of Soc. of Nav. Arch., Japan, Vol. 122, December 1967.
(12) N.H. Norrbin, "On the added resistance due to steering on a straight course", 13th Int. Towing Tank Conf., Berlin/Hamburg, 1972.
(13) M. Blanke, "Ship propulsion losses related to automatic steering and prime mover control", PhD thesis, Technical University of Denmark, 1981.
(14) C.G. Kallström, "Identification and adaptive control applied to ship steering", PhD thesis, Lund Institute of Technology, 1979.
(15) C.C. Lim and W. Forsythe, "Autopilot for ship design", Proc IEE, Vol. 130, Pt D, No. 6, Nov. 1983.
(16) D. Clarke, "Development of a cost function for autopilot optimisation", Proc Symp. on Ship Steering Automatic Control, Genoa, Italy, June 1980.

(17) R.E. Reid, et al, "Comparison of automatic steering performance of VLCC in a seaway resulting from application of LQG and classical control system design techniques", Proc 6th Ship Control System Symposium, Ottawa, Canada, Oct 1981.

(18) M.J. Grimble, "Controllers for LQG self-tuning application with coloured measurement noise and dynamic costing", Research Report No ICU/42/Apr 84, Strathclyde University, Scotland.

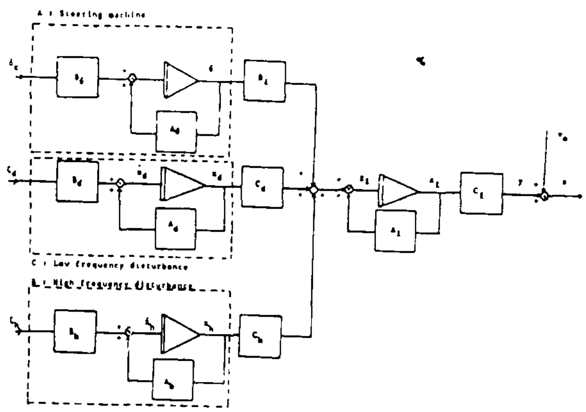


Figure 1 The wave model as an input disturbance

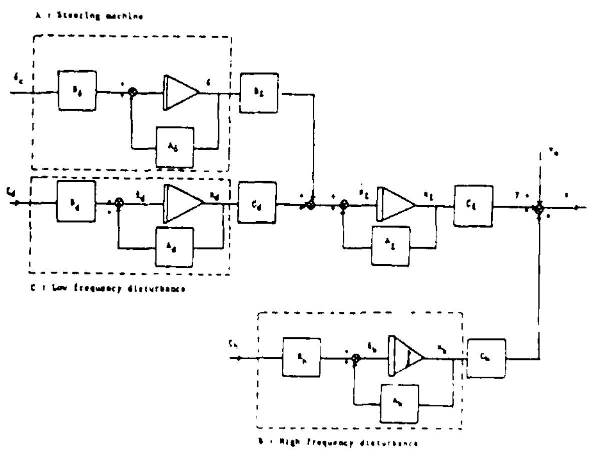
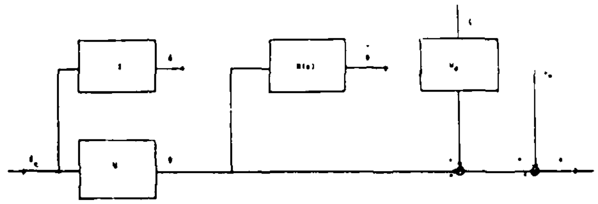


Figure 2 The wave model as an output disturbance



W : The ship transfer function

W_d: Wave model

v_ϕ: Measurement noise

ξ : White noise

Φ : The augmented state

Figure 3 : The representation of ship model and the augmented state in the frequency domain

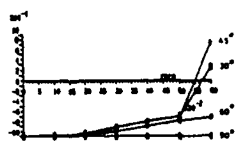


Figure 4 Variation of $\cos(\theta_r - \theta_d)$ with frequency and encounter angle

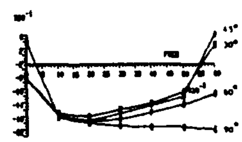


Figure 5 Variation of Φ_d with frequency and encounter angle

ONE APPROACH TO THE ON-BOARD WEATHER ROUTING AND ITS COUPLING TO THE OPTIMUM STEERING CONTROL

by Hideyuki Kanamaru
Mitsubishi Heavy Industries, Limited

ABSTRACT

It is universally acknowledged that the weather routing does contribute to achivement of the safety and economy of the voyage. It has already been put to practical service of late by some organs on land utilizing a huge volume of weather data and the satellite communication techniques. Yet it still is the captain who is charged with final determination of the ship's routes.

Introduced in this paper is one effective method to determine the ship's optimum sailing routes in the open seas. The process of route determination consists of the following two techniques: calculation of the shortest routes rounding about a hazardous region and selection of the safest and most economical route among them. Herein, the route efficiency index is discussed as a criterion of the routes and optimization method of the navigation schedule is presented. The on-board weather routing technique can be easily coupled to the optimum steering control through the route tracking technique.

INTRODUCTION

Amid the era of low growth in the national economy, strong demands for curtailment in costs of the shipping operations, not to mention the safety in the voyage, are being voiced in the shipping circles, and diversified countermeasures are being taken in respective fields of the hull, engine and electrical equipment in pursuit of ways and means for fuel-saving. Besides, as a part of these measures, efforts for rationalization of the ship's operations are steadily continued as well, and the optimum automatic steering system, which have hitherto accomplished a precursory role, is now about to enter a phase of its practical applications. On the other hand, much expectation has been focused on the utility of the weather routing in a fuel-saving voyage since more than ten years ago, nevertheless, a lack of accuracy in the meteorological and ocean weather forecasting as well as limitations of the communication means, etc. have so far hampered its application to practical uses. It was only recently that services utilizing the satellite communication, etc. by organs on land became available. In the first place, the final responsibility for operations of a ship is imposed on the captain, and, as for setting of the routes as well, it is an ordinary practice that the captain makes his decision thereof ultimately according to years' experience and intuition of his own, or based on supporting data transmitted from the land. This weather routing is a means for pursuing fuel-saving in a sphere completely different from that of the aforementioned optimum automatic steering, and it is requisite to make a systematic analysis of the voyage itself in order to achieve veritable rationalization of the ship's operations which comprise multiple phases as described hereabove.

In this report, an analysis of the voyage is firstly made from the viewpoint of the system engineering, succeeded by discussions on the fundamental techniques that serve for realization of the on-board weather routing, and an attempt is made lastly for systematization of respective element techniques involved in the processes from the route planning to the automatic steering, together with introduction of cases of their actual applications to the ship's operations.

SYSTEMATIC ANALYSIS OF NAVIGATION

The techniques for maneuvering the ship in the oceans can be classified broadly into two categories of the forward maneuvering and the preventive maneuvering. They are closely related mutually and form the voyage complementing reciprocally, as shown in Figure 1.

Forward Maneuvering

The forward maneuvering is a maneuvering operation for advancing the ship forward (namely, for having her approach the destination), and constitutes the very essence of the voyage. Another word, it has as its prime objective to realize a safe as well as economical voyage in the course of the ship's daily operations. This forward maneuvering can be conceived to have a hierarchical structure as described below.

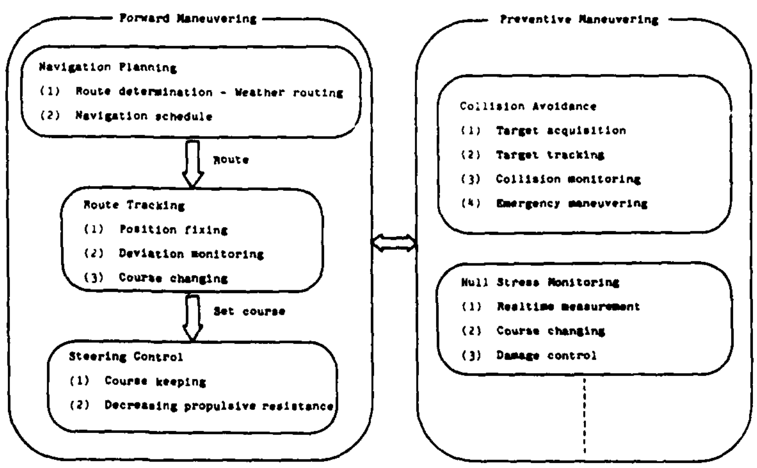


Figure 1. Two aspects of maneuvering

Navigation Planning. The navigation planning is an operation in which the routes up to the destination are set and a plan for navigation along these routes is drawn up prior to departure from a port, or thereafter depending on necessity. In the navigation planning, safest and most economical routes and navigation schedules are schemed off-line under a collective judgement formed on the basis of various data and information such as expected times of departure from, and arrival at, ports, pilot chart data, meteorological and ocean weather information or the ship's cargo load conditions. etc.

Route Tracking. The route tracking is an operation in which the courses and main engine output are controlled for achievement of faithful execution of tracking of the routes and navigation schedules preset in the navigation planning. A new course for tracking the routes economically is set upon arrival at each waypoint or upon deviating from a preset route. Besides, an on-line control of the main engine is carried out in order to assure the ship's arrival at each waypoint on schedule.

Steering Control. The steering control is an operation in which steering of the ship is exercised so as to maintain the courses set in the route tracking operation. The steering control contributes to realization of fuel-saving navigation of the ship, being performed in real time for the purpose of minimizing the propulsive resistance, adapting itself to the main engine output and sea conditions.

These three operations described hereabove constitute a hierarchical structure in which the lower-stratum operation functions according to the output data of the upper-stratum operation, and, in addition, each stratum is independently capable of achieving fuel-saving effects. Shown in Figure 2 is a functional block diagram of the forward maneuvering, while Table 1 indicates outlines and objects of the functions on each stratum together with means of their realization.

Table 1. Hierarchical structure and functional analysis of the ship maneuvering

Stratum	Function	Economy	Safety	Manpower-saving
Navigation planning (weather routing)	• Setting of appropriate routes from the viewpoint of economy(wide sea area) and safety(narrow sea area)	• Great circle sailing • Weather routing • Global route correction • Route analysis /evaluation	• Prevention of stranding • Weather routing	• Automatic search of the optimum route
Route tracking	• Setting of appropriate courses for tracking the routes preset in the navigation planning	• Optimum route selection • Utilization of tidal current data	• Recovery to the route • Confirmation of course changing operation • Local route correction • Utilization of tidal current data	• Automatic watch • Automatic judgement of hazards • Automatic hazard avoiding maneuvering
Optimum steering control	• Appropriate steering control to keep the desired course set by route tracking	• Optimum steering • Improvement of the steering gear	• Course keeping performance	

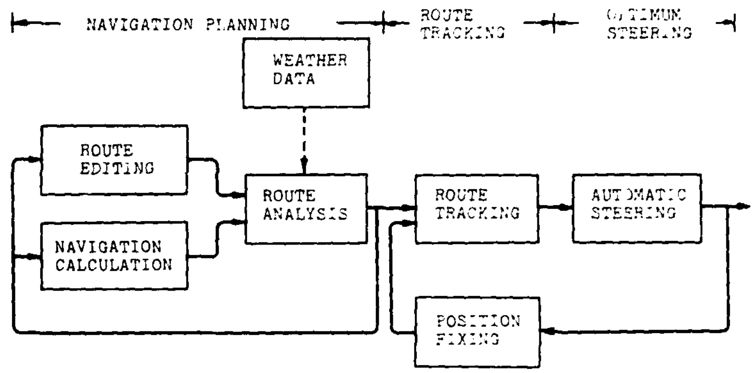


Figure 2. Function diagram

Preventive Maneuvering

In contract to the forward maneuvering which is closely related with the operation for advancing the ship forward, the preventive maneuvering is related with non-daily maneuvering such as an emergency and schedule adjustment operations. Its typical example is operation of the collision avoidance system (ARPA), which undertakes preventive functions for avoiding hazards rather than serving for directly advancing the ship forward. Needless to say, a system of this type is indispensable for assuring a safe voyage.

ON-BOARD WEATHER ROUTING

The principal technique of the on-board weather routing lies in the point of setting safe and economical routes up to the destination in due consideration of the global meteorological and ocean weather. At the organs on land which are currently providing services for the weather routing, meteorological forecasting is processed using a large-sized computer on the basis of relevant data accumulated over past decades, and specialists having qualification as a ship's captain are carrying out final route setting. While up to 72 hours ahead is said to be the maximum limit in the present technique for meteorological forecasting even using such a large-sized computer, the actual picture is that the captain determines the routes aboard the ship in consideration of the actual situations from a collective viewpoint by resorting to his years' experience and intuition or on the basis of rough data of pilot charts as well as weather routing information, etc. supplied in services from the land. In such a case, it is a most general practice to ensure safety in the voyage by avoiding hazards, if any, in the first place, then, to select economical routes. In conformity with the conception described above, the on-board weather routing technique proposed herein consists of two techniques of calculation of the optimum roundabout routes and the route analysis.

Roundabout Routing

The optimum roundabout sailing is a new and unprecedented sailing, and is defined as a "sailing along the route which links the point of departure and the destination in shortest distance, while avoiding designated hazardous region". The hazardous region is a rectangular sea area, on the Mercator's chart, demarcated respectively by two each of the meridional and equal latitudinal lines, and can be given independently from the limit latitude in the collected great circle sailing. Not only stormy weather but also shallow waters and archipelargos, naval maneuvers areas, etc. can be set freely as such. Shown in Figure 3 is an example of the optimum roundabout routes.

The most fundamental technique of the optimum roundabout route calculation lies in the point of processing calculations of the great circle route and collected great circle route as a package and forming them into a black box. Namely, a shortest route under a given condition is always calculated by means of putting into a package a series of processing for automatically judging the input data (positions of the point of departure/destination and, if required, the limit latitude) and obtaining the great circle route or collected great circle route as the output thereof, eliminating eventually necessity for being conscious of the difference between these two routes. The routes so obtained as the output from this package shall be specially referred to hereafter as "expanded collected great circle route".

The optimum roundabout route is determined by linking the expanded collected great circle routes, and the argorithm of this calculation is as stated below.

<u>Position Check of Departure Point and Destination.</u> If either of the point of departure or the destination is located on the higher side than the limit latitude or inside a hazardous region, it is treated as an error and relevant calculation is interrupted.

<u>Hazardous Region Crossing Check.</u> The expanded collected great circle route is calculated, and conditions of its crossing, if any, with a hazardous region are checked. If there exists no such crossing, the expanded great circle route is deemed as the optimum roundabout route. If crossed, however, processing transfers to search of a hazardous region roundabout route as described succeedingly.

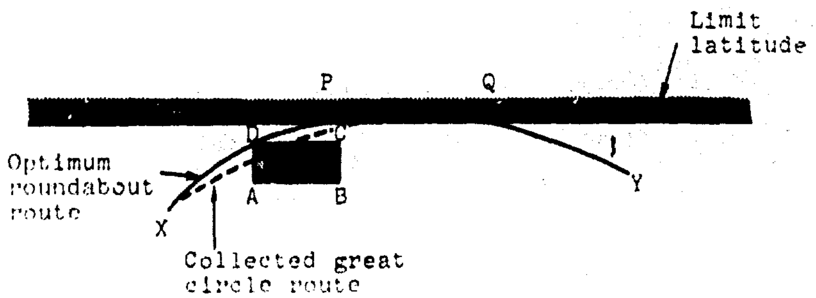


Figure 3. An example of the optimum roundabout route

Search of Roundabout Routes. Search of the hazardous region roundabout routes is performed as classified into the patterns shown in Table 2. Shown respectively as classified into patterns are, in the horizontal axis, "the positional relation between the point of departure/destination and the hazardous region" and, in the vertical axis, "the positional relation between the hazardous region and limit latitude", and search techniques for their combination are illustrated. Among the search methods of Case 1., search for the high latitude route is a system for determining the optimum roundabout route on the higher latitude side of the hazardous region, and its procedures are shown in Table 3. Further, search for the low latitude route consists of a system for determining the optimum roundabout route on the lower latitude side of the hazardous region similarily, procedures of which are as shown in the Table 4. The shorter one out of these two of the northern and southern routes is finally selected as the optimum roundabout route.

On the other hand, the east meridian route search and west meridian route search among the searching methods of Case 2. are meant for seeking the optimum roundabout routes respectively in contact with the east side and west side of the hazardous region, and, in this case, also, the shorter one out of these two is finally determined as the optimum roundabout route. Calculation results of the optimum roundabout route are shown in Figure 4.

The optimum roundabout route so determined is a route optimized only in conformity with the criterion of the "distance minimization" under the given condition for avoiding the hazardous region, and is devoid of requirements of the weather routing in which the meteorological and ocean weather conditions are to be put into account. This is because of placing its footing primarily on the principle of excluding in advance any route, which is already known to contain a large hazard, from the objects of the optimum route search, and this implies that priority is given to the safety rather than the economy.

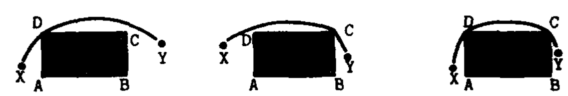
Figure 4. Calculation results of the optimum roundabout route

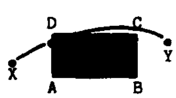
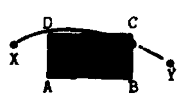
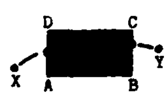
Table 2. Patternization of optimum roundabout route searching

Positional relation of the point of departure, destination and hazardous region / Positional relation of the hazardous region and limit latitude	(Case 1) In case either (both) of the point of departure or(and) destination is (are) located outside the meridian line on the boundary of the hazardous region.	In case both of the point of departure and destination are located between the meridian lines on the boundary of the hazardous region. (Case 2)	(Case 3)	(Case 4)
(Including a case when there is not the limit latitude.)	1 High latitude route search 2 Low latitude route search	1 East maridian line route search 2 West maridian line route search	1 Calculation of the expanded collected great circle route (Limit latitude:AB)	In this case, the route is determined upon checking of crossing condition with hazardous region.
	1 Low latitude route search	In this case, route calculation is rejected upon checking the position of the point of departure and destination.	1 Calculation of the expanded collected great circle route (Limit latitude:AB)	In this case, route calculation is rejected upon checking the point of departure and destination.
	In this case, route calculation is rejected upon checking crossing with hazardous region.			

Table 3. Highlatitude route search

Patterns of the optimum roundabout route / Pattern of the optimum roundabout route search process



Pattern of crossing with the hazardous region	Searched route / Shortest route	XDY	XCY	XDCY
1 	(a) XDY	◎	×	△
2 	(b) XCY	×	◎	△
3 	(a) XDY	◎	×	△
	(b) XCY	⊗	◎	△
	(c) XDCY	⊗	⊗	◎

The left table shows conditions of each searched route in the process of determining the optimum roundabout route.

◎ : Shortest route under a given condition

△ : It exists as the route but is not shortest. (It requires no trial.)

× : It does not constitute a route under a given condition. (It requires no trial.)

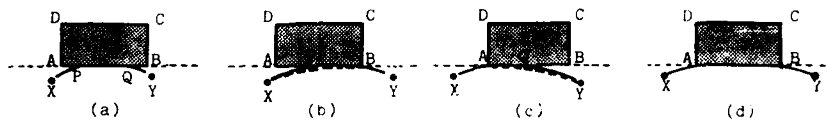
⊗ : It does not constitute a route under a given condition, but is required to be checked of its crossing with the hazardous region by a trial.

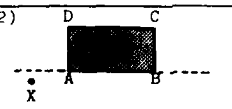
Remarks

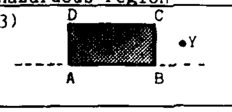
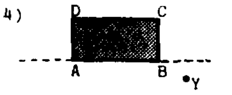
(Note 1) In the case of 3 in the above table, the prerequisite condition is that crossing with the hazardous region is confirmed in the sequence from XDY to XDCY.

(Note 2) The routes XDY, XCY and XDCY are all assembly of the expanded collected great circle routes.

Table 4. Low latitude route search



Pattern of the optimum roundabout route / Optimum roundabout route search process	Positional relation of the point of departure and hazardous region	Searched route / Shortest route	XPB(XB)	XAB
Point of departure side search Point of departure : X Point of temporary destination : B Limit latitude: AB Only (2)	(1) 	XAB	×	◎
	(2) 	XPB (XB)	◎	△
		XAB	⊗	◎

	Positional relation of the destination and hazardous region	Searched route / Shortest route	YQA(YA)	YBA
Destination side search Point of departure : Y point of temporary destination : A Limit latitude: AB Only (4)	(3) 	YBA	×	◎
	(4) 	YQA (YA)	◎	△
		YBA	⊗	◎

3. Formation of the optimum roundabout route

Destination side / Point of departure side	YQA (YA)	YBA
XPB (XB)	(a) XPQY	(b) XPBY (XBY)
XAB	(c) XAQY (XAY)	(d) XABY

(Note 1) Refer to Table ? for the symbols (◎ , △ , × , ⊗) in the left table.

(Note 2) The routes of XAB, XQA, etc. are all assembly of the expanded collected great circle routes and parallel sailing routes.

Route Analysis

The route analysis is a technique in which the economical aspects of the routes is analyzed upon putting into account the meteorological and ocean weather conditions, and is the pivotal function among those of the on-board weather routing. Namely, it is a technique which constitutes the basis for determining a route out of a group of the optimum roundabout routes, from which possibility of serious hazards, if any, has been excluded. A reasonable basis, namely, a criterion for evaluation of the routes, is necessary for determining the route.

Criterion for Route Analysis. Evaluation of the routes in the on-board weather routing shall be performed on the basis of the overall energy consumption from departure from a given port to arrival at a given destination, and not merely based on fuel consumption per a unit of time, and the criterion therefor is defined as follows.

$$J_R = (1-\mu)\int_{t_d}^{t_a} P dt + \mu\int_{t_d}^{t_a} \Delta P dt \qquad (1)$$

Where, P : Propulsive horsepower (engine output) in completely calm sea
ΔP: Propulsive horsepower loss by waves
t_d: Time of port departure
t_a: Time of port arrival
μ : Weighting coefficient ($0 \leq \mu \leq 1$)

The propulsive horsepower is expressed, as known well, as a sum of frictional resistance and wave making resistance.

$$P = [C_W + (1+\kappa)C_f]\frac{\rho}{2}\cdot U^3 \cdot \nabla a^{2/3} \qquad (2)$$

Where, C_W: Coefficient of wave making resistance
C_f: Coefficient of frictional resistance
κ : Form factor
∇a: Sea water density
U : Ship's speed

On the other hand, loss of propulsive horsepower by the waves is defined in the following formula, using the speed characteristics both in completely calm sea and in the waves.

$$\Delta P = P - [C_W + (1+\kappa)C_f]\frac{\rho}{2}\cdot(\alpha U)^3 \cdot \nabla a^{2/3} \qquad (3)$$

where, α is the natural speed reduction rate in the waves, and the 2nd term in the right side member of Formula (3) is the "propulsive horsepower on the assumption of the ship sailing in completely calm sea with ship's speed reduced by the waves". The criterion is given

$$J_R = \int_{t_d}^{t_a} [(1-\mu)+\mu(1-\alpha^3)]\cdot[C_W+(1+k)C_F]\frac{1}{2}\rho U^3\cdot\nabla^{2/3}\, dt \qquad (4)$$

As ΔP of Formula (3) is in a positive correlation with additional propulsive resistance received from the waves, namely, the external force of waves which the hull undergoes, the criterion J_R of Formula (4) includes criterion on the safety of the route.

Speed Performance in Sea Condition. Several researches have been made pertaining to the speed reduction performance in the waves[3]-[5], and Hagiwara and Nakishima have proposed the following approximate expression taking up the container as an example.[9] Namely, as regards the natural speed reduction performance;

$$U_w = U_O(P) - m(P)\cdot f(h)\cdot g(\psi) \qquad (5)$$

$$U_O(P) = 0.905P^{1/3}$$

$$m(P) = 1.4 - 2.8 \times 10^{-5}\cdot P$$

$$f(h) = 12.0[1.0 - \exp(-3.2 \times 10^{-3}h^{2.8})]$$

$$g(\psi) = 0.8 \exp(-4.1 \times 10^{-5}\psi^{2.4}) + 0.2$$

where, U_w: Ship's speed in the waves (KT)
P : Engine output (BHP)
h : Significant wave height (m)
ψ : Relative wave direction (deg)

and, the sailing limit speed in the waves is;

$$U_\ell = \exp[0.13\{q(\psi) - h\}^{1.6}] + r(\psi) \qquad (6)$$

$$q(\psi) = 12.0 + 1.4 \times 10^{-4}\psi^{2.3}$$

$$r(\psi) = 7.0 + 4.0 \times 10^{-4}\psi^{2.3}$$

Formulas (5) and (6) being approximate expressions of the speed reduction performance pertaining to a model ship, parameters and/or speed reduction performance curves corresponding to each ship are required in general. At this time, the natural speed reduction rate α is given in the following formula.

$$\alpha = \frac{U_W}{U_O(P)} \qquad (7)$$

On the other hand, when there is a tidal current, the ship shall advance on the route riding the tidal current for route tracking.

In this case, the ground speed is [illegible] according to Figure 5.

$$V = \sqrt{(\alpha U)^2 - (v \sin\theta)^2} + v \cos\theta \qquad (8)$$

where, V: Ground speed
v: Tidal current speed
θ: Difference between the bearing of the tidal current and the bearing of the route

Besides, the set course is required to be deviated from the bearing of the route by the following value.

$$\phi = \sin^{-1} \frac{v \sin\theta}{\alpha U} \qquad (9)$$

The tidal current does not directly influence the propulsive horsepower like the waves do, but it becomes related with the criterion J_R of Formula (4) in the point that the navigation time up to the destination varies depending on utilization of the tidal current.

Navigation Scheduling

The navigation scheduling is a technique for optimizing a method to guide the ship to the destination along given routes. In the ordinary ocean navigation, rescheduling of set course and/or ship's speed is done intermittently at the time of arrival at a waypoint or relief of watches. Accordingly, the optimization of the navigation schedule shall be done, as shown in Figure 6, on the basis of a route between two waypoints (route segment) as a unit therefor. At this time, putting respectively additive notes of "0" to the place of departure, "N" to the destination and "i" to each waypoint ($i = 0 \sim N$), and expressing the route between No.i and No.(i + 1) as the "route segment i", then, the criterion J_R of Formula (4) can be expressed as follows.

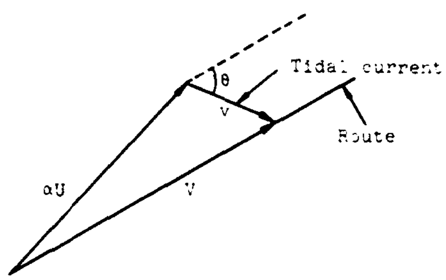


Figure 5. Tidal current correction

$$J_R = \sum_{i=0}^{N-1} [(1-\mu)+\mu(1-\alpha_i^3)] \cdot [C_W + (1+ \text{[illegible]}) \cdot U_i^3 \cdot \text{[illegible]}] \cdot \frac{\ell_i}{\sqrt{(\alpha_i U_i)^2-(v_i \sin\theta_i)^2}+v_i \cos\theta_i} \qquad (10)$$

where, U_i: Ship's speed at the time when the route segment i is in completely calm sea

α_i: Average natural speed reduction rate in the route segment i

v_i: Average tidal current speed in the route segment i

θ_i: Difference between the bearing of the average tidal current and the bearing of the route in the route segment i

ℓ_i: Distance of the route segment i

Here, if the expected port departure time t_d and the expected port arrival time t_a are specified, the navigation time T is determined.

$$T = t_a - t_d \qquad (11)$$

This constitutes a binding condition in the optimization of the navigation schedule. Namely,

$$\sum_{i=0}^{N-1} \frac{\ell_i}{\sqrt{(\alpha_i U_i)^2-(v_i \sin\theta_i)^2} + v_i \cos\theta_i} = T \qquad (12)$$

Ultimately, the optimization of the planning on the ship's speed is performed by seeking U_i ($i = 0 \ldots N-1$) which minimizes the criterion of Formula (10) under the binding condition of Formula (12). Using the Lagrange's method of indeterminate coefficients,

$$\hat{J}_R = J_R + \lambda\left[\sum_{i=0}^{N-1} \frac{\ell_i}{\sqrt{(\alpha_i U_i)^2-(v_i \sin\theta_i)^2} + v_i \cos\theta_i} - T\right] \qquad (13)$$

where, λ is an additive variable.

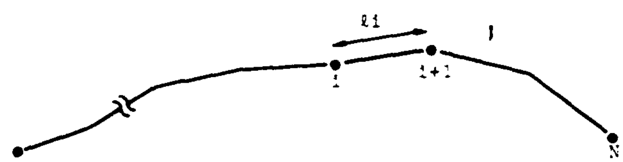


Figure 6. Route segment

By setting the following formulas,

$$\frac{\partial \hat{J}_R}{\partial U_i} = 0 \quad (i = 0 \sim n-1) \tag{14}$$

$$\frac{\partial \hat{J}_R}{\partial \lambda} = 0 \tag{15}$$

following simultaneous equations can be formulated.

$$2\alpha_i^2 U_i^3 + 3v_i\cos\theta_i U_i\left(\sqrt{(\alpha_i U_i)^2-(v_i\sin\theta_i)^2} + v_i\cos\theta_i\right) - 3v_i^2 U_i - \frac{\alpha_i^2 \lambda}{[C_w + (1+\kappa)C_f]\frac{\rho}{2}\,\nabla a^{2/3}(1-\mu\alpha_i^3)} = 0 \quad (i = 0 \sim n-1) \tag{16}$$

$$\sum_{i=0}^{N-1} \frac{\ell_i}{\sqrt{(\alpha_i U_i)^2-(v_i\sin\theta_i)^2} + v_i\cos\theta_i} = T \tag{12}$$

By solving this equation relating to $U_i (i = 0 \sim N - 1)$, an optimum planned value of the ship's speed is obtained. By the way, if the ship's speed U_i changes, the relative wave direction changes owing to the tidal current correction (9). Furthermore, as the engine output P also changes, the natural speed reduction rate changes correspondingly. Accordingly, it is necessary for planning the optimum ship speed to solve Formulas (12) and (16) simultaneously, and, further, to let the solution converge using Formulas (5) and (7). Particularly in case there is no tidal current, Formulas (12) and (13) can be solved easily, and the optimum ship's speed is determined as follows.

$$U_i = \frac{1}{TA_i} \sum_{i=0}^{N-1} \frac{A_i \ell_i}{\alpha_i} \tag{17}$$

$$A_i = \sqrt[3]{2(1-\mu\alpha_i^3)[C_w + (1+\kappa)C_f]\frac{\rho}{2}\cdot\nabla a^{2/3}}$$

COUPLING TO OPTIMUM STEERING CONTROL

Various kinds of researches have been performed up to now as regards the optimum steering aiming at fuel-saving[4]-[8], but it is is only recently that relevant technology is realized in the form of actual equipment. This features improvement in the steering efficiency by appropriate steering control and consequential contribution to fuel-saving, and its criterion can be expressed by the following formula on the basis of the principal particulars of hull, propulsive performance and steering maneuverability performance, etc.[5]

$$J_s = \lim_{T\to\infty} \frac{1}{T} \int_0^T (x'Qx + 2x'Su + u'Ru)dt \tag{18}$$

where, "'" of Formula (18) means a transposed matrix.

x and u are vectors of state variables respectively indicating the maneuvering conditions of the ship and rudder angle.

$$x = \begin{pmatrix} \Delta\psi \\ r \end{pmatrix} \tag{19}$$

$$u = \delta$$

Besides, Q, S, R are weighting coefficient matrices, and are determined uniquely by the principal particulars of hull and maneuvering performance of the ship.

$$Q = \begin{pmatrix} \frac{1}{2} & \frac{a}{2U} \\ \frac{a}{2U} & \frac{a}{U^2}\left[\frac{a}{2} - \frac{m_o + m_y}{\{C_w + (1+\kappa)C_f\}\frac{\rho}{2}\nabla a^{2/3}}\right] \end{pmatrix} \tag{20}$$

$$S = \begin{pmatrix} \frac{b}{2U} \\ \frac{b}{2U^2}\left[a - \frac{m_o + m_y}{\{C_w + (1+\kappa)C_f\}\frac{\rho}{2}\cdot\nabla a^{2/3}}\right] \end{pmatrix}$$

$$R = \frac{b^2}{2U^2} + \frac{\eta(1-\omega)^2(1+3.6S^{1.5})}{\{C_w + (1+\kappa)C_f\}\frac{\rho}{2}\cdot\nabla a^{2/3}} \cdot \frac{A_R\lambda}{\lambda + 2.2}$$

here,

$$a = \frac{\hat{K}}{K}\left(1 - \frac{\hat{T}_3 - T_3}{T_1 + T_2 - T_3}\right)$$

$$b = \hat{K}\cdot\frac{\hat{T}_3 - T_3}{T_1 + T_2 - T_3}$$

On the other hand, a K-T model as the ship's steering maneuver model is assumed, and is expressed by the following differential equation.

$$\dot{x} = Ax + Bu \tag{21}$$

A and B are coefficient matrices, and are expressed respectively as follows.

$$A = \begin{pmatrix} 0 & 1 \\ 0 & -\frac{1}{T} \end{pmatrix} \quad (22)$$

$$B = \begin{pmatrix} 0 \\ -\frac{K}{T} \end{pmatrix}$$

Here, the optimum control law is of the following feedback control.

$$u = -R^{-1} B' P x \quad (23)$$

P is a solution of the following Riccati's differential equation.

$$PBR^{-1}B'P - PA - A'P - Q = 0 \quad (24)$$

As described hereabove, the optimum steering is an operation for pursuing fuel-saving in a sphere completely different from that of the weather routing, and it is possible to expand the fuel-saving effect further by combining these two.

The on-board weather routing and optimum steering can be coupled together through the route tracking function stated above. Namely, an optimum weather route is determined by the on-board weather routing, and the course along which the ship is to advance are determined by means of the route tracking function. And, the optimum steering function works so as to keep the courses most economically. By organic combination of the series of functions, an integrated automatic maneuvering covering the processes from the navigation planning to maneuvering becomes possible.

A systematic diagram of the total system in which all the functions from the on-board weather routing to the optimum steering is integrated is shown in Figure 7. This system is divided into the optimum navigation planning system and the optimum steering control system, and the route tracking function is contained in the optimum navigation planning system. As the meteorological and ocean weather data, data of the world's oceans contained in the pilot charts are incorporated therein as classified for each month, but it is possible to use these data as modified to cope with the actual conditions. On the other hand, it is so designed as to be capable of receiving precise positional data from the satellite navigation system as an external equipment, and is also equipped with a single-loop steering gear, in which a servo motor is adopted, resulting in success in further improvement in the fuel-saving effects. This system was installed on a bulk carrier of 207,000 DWT, and is still continuing smooth operation at present.

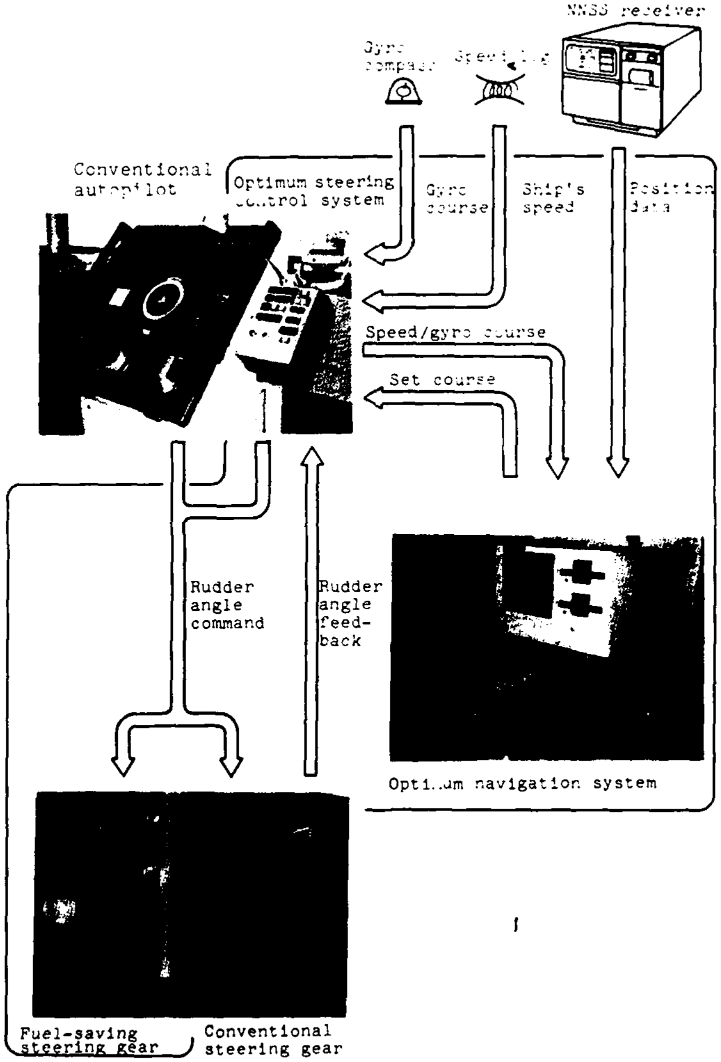


Figure 7. Integrated navigation system

CONCLUSION

Development of new navigational equipments is currently accelerated as stimulated by strong needs for improvement in the economy and manpower-saving, etc. in the shipping operations. However, it is necessary not only to aim at "electronization" of the conventional sailing techniques but also to develop new techniques themselves. Besides, it is required, at the same time, to analyze the navigation from the viewpoint of the system engineering, and systematically build up the element techniques for supporting the navigation in such a manner as will be in line with the form as it should be.

Introduced in this paper is an on-board weather routing technique newly developed upon designing systematization of the maneuvering techniques from the navigation planning to steering as an initial stepping stone thereof. This technique is formed on two fundamental techniques of the optimum roundabout routing and route analysis, and can be coupled organically with the steering control through the route tracking. Statistic data contained in the pilot charts have been used as the meteorological and ocean weather data, however, if it becomes possible aboard to obtain instantaneously the meteorological and marine ocean data by consolidation of the marine satellite system, diffusion of the remote sensing technique or satellite communication technique in the near future, then, we can apply the technique introduced herein as it is only by replacing the data of the pilot charts by the data which can be collected in real time. In addition, an attempt to organically integrate the whole operations from the navigation planning to the optimum steering still remains merely as a step towards the systematization of the bridge, and it constitutes one of the element techniques for the next overall integration to come. It is sincerely anticipated that the various techniques and systems herein introduced will serve as the nuclear techniques to contribute to achievement of the aforesaid objective.

REFERENCE

(1) S. Nakamura and S. Naito, "Nominal Speed Loss and Propulsive Performance of a Ship in Waves", Journal of the Kansai Society of Naval Architects, Japan, No. 166, 1977, p. 25.

(2) S. Naito, S. Nakamura and S. Hara, "On the Prediction of Speed Loss of a Ship in Waves", Journal of the Society of Naval Architects of Japan, Vol. 146, 1979, p. 147.

(3) H. Hagiwara and T. Makishima, "A study on the Optimum Ship Routes", the Journal of Japan Institute of Navigation, Vol. 62, 1980, p. 77.

(4) K. Ohtsu, M. Horigome and G. Kitagawa, "A Robust Autopilot System against the Various Sea Conditions", Ship Operation Automation, 3rd International Symposium, 1979, p.118.

(5) H. Kanamaru and T. Sato, "Adaptive Autopilot System with Minimum Propulsive Energy Consumption", Ship Operation Automation, 3rd International Symposium, 1979, p. 124.

(6) J. van Amerongen, "Experiences with a Digital Model Reference Adaptive Autopilot", Ship Operation Automation, 3rd International Symposium, 1979, p. 134.

(7) D.R. Broome, "The effect of Autopilot Tuning on Ship Sea Keeping Performance", Ship Operation Automation, 4th International Symposium, 1982, p. 37.

(8) T. Koyama, "On the Adaptive Automatic Steering System of Ships at Sea", Journal of the Society of Naval Architects of Japan, Vol. 122, 1967, p. 18.

(9) K. Nomoto and T. Motoyama, "Loss of Propulsive Power Caused by yawing with Particular Reference to Automatic Steering", Journal of the Society of Naval Architects of Japan, Vol. 120, 1966, p. 71.

(10) J.K. Zuidweg, "Course-keeping as a Special Case of Track-keeping", Ship Operation Automation, 4th International Symposium, 1982, p. 65.

(11) J. Van Amerongen, "An Adaptive Autopilot for Track-keeping", Ship Operation Automation, 3rd International Symposium, 1979, p. 99.

(12) A.W. Brink, G. Bass, A. Tiano and E. Volta, "Adaptive Automatic Course-keeping Control of a Super-tanker and a Container Ship -- A Simulation Study", Proceedings 5th Ship Control Systems Symposium, Vol. 4, 1978, p. 4-1.

(13) "Total Fuel-saving Navigation System Developed", ZOSEN, Vol. 27, No. 5 , 1982, p. 20.

(14) "Test Report on Total Fuel-saving Navigation System Developed by Mitsubishi H.I. and JAMDA", ZOSEN, Vol. 28, No. 5, 1983, p. 34.

FAULT DETECTION: BAD DATA, HOW TO FIND IT AND HOW TO HANDLE IT

JOHN R. WARE

ORI, Inc.
1375 Piccard Drive
Rockville, MD 20850
USA

ABSTRACT

With the large volume of data being transmitted in current and proposed monitoring and control systems, it is important to detect invalid data and failed transmission lines. Detection of invalid data is the first step in the construction of a fault-tolerant system and this paper defines several basic terms with regard to fault and failure detection and proceeds to discuss many of the methods that have been used or suggested. The advantages and disadvantages of these methods are presented, together with several suggestions for control processing during brief periods in which valid data may not be available.

INTRODUCTION

With the increased use of complex automation systems in ship control, there has been a corresponding increase in the need for high volumes of data transmission. Data are not only used for control, but also for Performance Monitoring and Fault Location (PM/FL) systems, operator displays, data recording systems, and so forth. With the increased level of automation, we have typically reduced the number of personnel who are in charge of ships' systems and have, therefore, become more dependent upon the automatic system's performance. Thus, we have a two-fold problem: the large volume of data being transmitted means it is more likely that some data are invalid (either due to a failure in a sensing system, improper A/D conversion, failure in the transmission system, or failure in the receiving system); and it is less likely that a human will be able to detect the error.

In this paper, we will discuss the important topic of invalid (or "bad") data, how to detect it, and what can be done about it in the control system. We will start with a number of definitions that, while somewhat arbitrary, will serve to clarify the ensuing discussion. It would appear that the lack of clear-cut definitions has severely hampered the transfer of information among various researchers and has currently caused problems in several failure detection systems used on shipboard. Using the definitions, the types of possible faults will be described, together with the possible outcomes of a fault detection scheme and their consequences. A number of fault detection methods will be described, beginning with the simple magnitude tests and proceeding to more complex methods based, typically, on optimal control and estimation techniques. Following this, we will present a concise description of the failure determination problem; that is, how much bad data will we accept from a particular source before declaring that source to be permanently failed? Finally, we will discuss "what to do until the doctor comes." That is, given that we have detected some bad data, and there is no alternative source, how should we best handle the situation?

It is important to emphasize several points with regard to this paper. First, the subject is the detection of "bad" data; that is, data that are not representative of the physical state of the system being observed. Therefore, failures in physical systems being controlled or monitored, such as would result from a rudder jam or lubricating oil overheating, are not considered "faults" in this context as long as the data line reports the situation correctly. Second, in order to focus on a the topic of data fault detection, fault isolation techniques are not described (although several of the methods mentioned inherently contain fault isolation information). Third, and last, this work, should not be taken as a comprehensive survey of all available methods, but as a representative survey based on the author's experience and available material. Much of the information presented here has not been published in the open literature, but is actually being used in ship-borne control and monitoring systems. Where open literature citations are available, references are made; but a lack of reference is not meant to imply that the method described is not being used at sea. In fact, in general, referenced methods are <u>not</u> being used at sea (to my knowledge), while unreferenced methods are.

DEFINITIONS

One of the reasons that we seem to have so many difficulties in regard to the whole question of faults, failures, bad data, and so forth is a lack of clear definitions of the items under discussion. In one circumstance, this has led to the ludicrous situation that the only way in which a particular failure detection and data handling system would function is for one part of the system to assume that, whenever the fault detection system discovered an error, it was wrong and the data were actually correct. Let us, therefore, consider the following working definitions:

<u>DATA LINE:</u> A data line is the entire sequence of devices and operations that operate on a single datum from the point at which it is sensed to the point at which the datum is in a form and location suitable for evaluation. Thus, a data line may include the sensor, local signal processing, A/D conversion, data transmission, multiplexing, or any other operations needed to present the datum to the control and/or monitoring system.

<u>FALSE DATA</u>: Data are <u>false</u> when they are no longer representative of the state of the <u>system</u> being used.

In this sense, almost all data are false to some degree due to normal system and measurement noise, digital quantization, and such. This leads to the more useful definition:

INVALID DATA: Data are invalid when they contain insufficient information to support the purpose for which they are being gathered.

It is important to note the subjectivity of this definition, which is a vital part of the definition. If data are being gathered for use by two different functions, one may feel that, under certain circumstances, the data are invalid and cannot be used, while the other may feel that the data are sufficiently accurate for its needs. In a particular example, both the ship's defensive weapons systems and the automatic control system required heading information from the inertial navigation system. Under a certain unique combination of circumstances, the data source developed a small oscillation at a frequency approximately 20 times that of the closed loop heading control system. Thus, internal filtering in the automatic heading loop was able to attenuate this oscillation with almost no degradation to heading control. However, the defensive weapons system which performed the data checking declared the heading data invalid (for its purposes), and caused the heading control system to default to manual control needlessly.

It is important to distinguish between the concepts of "failures" and "faults." In the U.S., these are often used interchangeably, even within the Naval community. In order to attempt to resolve this ambiguity, we have resorted to the dictionary to provide basic guidelines. Within the context of data transmission, the following appears appropriate:

FAULT: A data line has faulted when a single datum is invalid. That is, a fault is the occurrence of one non-useful piece of information.

The important point concerning fault is that it may be a temporary situation that may resolve itself. This is as opposed to a failure:

FAILURE: A data line has failed when it continually supplies invalid data and can no longer support the purpose for which the data are being transmitted.

In most instances in which data are being transmitted for purposes related to ship control, a single fault is not considered to be sufficiently serious to label the data line as having failed. Faults can occur for a number of reasons that do not necessarily imply that the data line has failed. For instance, a burst of electrical noise may cause a momentary disruption of the rearrangement of a single bit pattern in the data string. The problem of when to declare a failure is a difficult one which will be discussed later.

The distinction between "fault" and "failure" is not important, but may be critical depending upon the detection schemes that may be employed. As will be seen later, many of the advanced techniques will only work properly when the data line has only two possible states: functioning or failed. If a line experiences several faults and then recovers, the schemes either do not detect the data as being invalid or declare the line as failed prematurely. This can be a major flaw and one which is usually not mentioned in the literature describing these methods.

FAULT TYPES AND FAULT DECISIONS

There are many ways in which a data line can produce a fault and these may be characterized by the effect they have on the sensed data. These include:

- o Failure to a maximum (or minimum) position; that is, a hard over failure.
- o A bias failure in which the sensed value suddenly develops a constant bias. This would occur, for example, if a bit "locked" in some element of a digital transmission.
- o An increasing bias, or drift error, as a sensor slowly degraded in accuracy.
- o A total "lock" failure in which all data remain at a constant value.
- o A random failure in which received data are almost totally noise with little or no signal content.

Normally, the types of failures that are most likely will drive the method of fault detection as, for economic reasons, the surest method of fault detection (redundancy) is not always economically feasible.

Whenever a fault detection system is implemented, it must continually make decisions with regard to the received data. As with most decisions, the possibility of error exists and the four possible outcomes of the decision process are as follows:

1. The decision is that the data are valid when the data are actually valid.
2. The decision is that the data are invalid when the data are actually invalid.
3. The decision is that the data are valid when the data are actually invalid.
2. The decision that the data are invalid when the data are actually valid (false alarm).

Two of the four possible outcomes are errors, each of which can have potentially serious consequences. The results of an error of accepting invalid data as being valid are fairly obvious in that loss of stability may occur, with the seriousness of the outcome dependent upon the application and the particular data sources.

The consequences of false alarms are not so immediate or apparent. However, if an excessive number of false alarms occur, then there may be a tendency on the part of operational personnel to ignore all alarms, even those that are valid, or to attempt to circumvent the fault detection logic in some manner. Both of these eventualities have occurred in practice.

Unfortunately, the two types of errors are not independent. Every fault detection scheme has some associated threshold which, if exceeded, results in a declaration of a fault. The more sensitive the scheme, that is, the lower the threshold, the more likely it is that false alarms will occur. On the other hand, if the threshold is increased to minimize the number of false alarms, then the possibility of declaring invalid data as being valid is increased. There are no hard and fast rules for making decisions in the design of fault detection systems or for selecting these thresholds (or, equivalently, the associated probabilities); and judgment must be tempered with experience and a knowledge of the consequences of each type of error. In a highly redundant system, the consequence of a false alarm is simply to lower the level of redundancy for whatever time it takes for human intervention to determine that a false alarm has actually occurred. This may be a negligible problem or a major headache, depending on the application.

FAULT DETECTION METHODS

Fault detection methods may be grouped roughly, and in approximately the order of increasing complexity, as follows:

- Magnitude tests
- Rate tests
- Prediction methods
- Comparison tests (line redundancy)
- Consistency tests (analytic redundancy)
- Miscellaneous advanced methods

The method, application, advantages, and disadvantages of each of these will be discussed briefly in the following sections.

Magnitude Tests

The simplest of the fault isolation tests is simply to compare the magnitude of the data with some known physical limit that cannot be exceeded. Common examples would be speed in excess of some value, say 100 knots, or a heading in excess of 360 degrees. While magnitude tests are not very discriminating (e.g., if a ship typically cruises at 6 knots but is capable of 35 knots, the magnitude tests must still be set in excess of 35 knots; thus allowing considerable error during the majority of operational time), they may be useful in non-complex systems when more elaborate testing is precluded due to lack of computational space or time. Typically, magnitude tests will only detect hard over failures or gross shifts in data line bias. However, these are not an insignificant class of faults and the simplicity of the magnitude test does not cause a computational burden. Magnitude tests may be used in conjunction with more elaborate methods which are not brought into play until triggered by a magnitude test failure.

Rate Tests

Rate tests are only slightly more complex than magnitude tests. They require computing the differences between two consecutive data samples and comparing this difference to a known maximum. For example, if the maximum turn rate of a ship were known to be 2.5 degrees per second and a difference of two consecutive samples was 0.3 degrees with a sampling rate of 10 times per second, then one of the data points must be in error. Rate tests are often used in conjunction with magnitude tests and can detect sudden hard over failures, bias failures if the bias is large enough, drift errors if the drift is fast enough, and some random failures. The price that is paid for this increase in capability is the corresponding increase in difficulty of determining the requisite rate limits which may require analysis or simulation and which may be functions of operational conditions, such as ship's speed. Additionally, because the act of differentiation is potentially "noisy", several rate values may need to be averaged (or the rate filtered). This has the double drawback of possibly masking some faults and of delaying the detection of others due to the time lag inherent in the averaging process. The alternative is to increase the rate limit to account for the possible noise and, correspondingly, to decrease the sensitivity of the test.

Prediction Methods

A technique that is closely related to rate testing, and which, therefore, shares some of its advantages and disadvantages, is to use past data to estimate what the current value of the data should be. Of the possible approaches, the simplest is to use the previous value of the data along with an estimate of the rate to estimate the current value by assuming an approximately constant rate over the time interval between samples. This procedure can be extended by including more past information, such as the past value of acceleration. In this simple form of the method, prediction is essentially the same as extrapolation using N past data values. When N = 2, rate only is used for the prediction; when N = 3, acceleration is used for prediction; and so on.

Prediction can involve more advanced procedures by using a recursive filtering procedure. In a relatively noise-free application, a recursive least squares predictor could be used. If the underlying system dynamics are linear and known, and if the measurement and process noise are known and meet the usual requirements, then a Kalman predictor can be used to provide the best (in the mean square sense) prediction. In this case, the prediction method would more properly fall under the general heading of "advanced methods", which will be discussed below.

The major advantage of prediction methods is that they are generally more sensitive then simple rate rests for determining sudden bias errors or random errors. Another potential advantage of prediction methods is that, should the current data be found to be invalid, then it may be possible to use the predicted value for the purposes of control or display generation. The major disadvantages are that they require more knowledge of the underlying dynamic process and the measurement noise in order to place acceptance bounds on the difference between the predicted and actual measurement; and they require greater computational and storage resources.

Comparison Tests

Comparison tests require that at least some portion of the data line be redundant so that at least two partially independent samples of the data are available. If the redundancy is dual, then a comparison of the two samples is made, and, if the difference exceeds predetermined bounds, a fault has been detected. Naturally, in dual redundant systems, the comparison cannot be used to determine which of the lines has faulted and some other method may be employed to ascertain which line is more likely to have failed. A dual redundant system was suggested by Deckert et al. (Reference 1) for the F-8 aircraft in conjunction with a detection and hypothesis testing procedure based on analytic redundancy (c.f., below). The advanced method was not employed until the dual system comparison indicated a need for more detailed failure detection and location.

Despite the basic problem of dual redundancy, that is, the inability to determine which data are invalid, such systems are frequently employed where cost and/or space are premiums and where continued functioning of the system is not imperative. Thus, dual redundancy is common in automatic ship control where a manual back-up is quickly available and can perform the task with comparable skill. However, in applications which are more critical, triple and even quad redundancy is employed. With the rapid rise in microprocessor technology and the parallel decrease in processing costs, highly redundant systems are no longer as prohibitive as they were only a few years ago.

When the data redundancy is more than dual, the comparison becomes a voting procedure in which majority rules (with the natural assumption that two simultaneous and equal faults is a highly unlikely occurrence). The complexity of highly redundant systems, which may include redundant control computers, leads to sophisticated software and hardware interactions for ensuring that all systems are equal and capable of performing the comparisons equally; and for planning data paths in the face of multiple hardware failures.

The presence of multiply redundant data lines provides the possibility of improving data accuracy by averaging the received data. An averaging procedure can reduce measurement noise by a factor of the reciprocal of the square root of the number of independent samples available. In a generalization of this approach, Broen has suggested a general linear combination of the samples, with weighting factors selected to minimize the error covariance (Reference 2).

Comparison schemes are, in principle, easy to implement if a single computer is used to evaluate the data. They are the only sure means of detecting virtually every type of failure, with the possible exception of small bias errors. Multiply redundant systems also have the desirable property of being able to continue to function by utilizing the valid data or, if the location of the failure can be isolated, by reconfiguring the data path. The major disadvantage of redundant systems is cost for single computer systems, and the complexity of software/hardware for multiple computer systems. When multiple computers are used, each may perform input and output comparisons involving all the other systems. This can significantly increase the hardware and/or software burden and, correspondingly, increase the probability of failure due to the increased parts count.

Consistency Tests

Consistency tests, often referred to as analytic redundancy tests, are attempts to take advantage of the functional relationships that may exist between available data. For example, a heading signal can be differentiated twice and the resulting acceleration compared with the acceleration that would result from the existing rudder angle, ship's speed, and lateral velocity, from the inertial guidance system using known ship's coefficients. Other forms of analytic redundancy often exist in machinery monitoring and control systems when sufficient information is being gathered. For instance, pump RPM, inlet pressure, outlet pressure, and flow rate are not independent quantities and their relationship, if known with sufficient accuracy, can form the basis for a consistency test.

The quality and sensitivity of a consistency test is highly dependent upon the quality of the available measurements (note the double differentiation described above) and of the level of knowledge of the dynamics being monitored and controlled. Because of this, tests employing analytic redundancy often employ complex filtering and error detection mechanisms, such as the sequential probability ratio test (SPRT) proposed by Deckert et al. (Reference 1). The complexity of these tests may preclude them from running continuously because of the computational burden imposed and they therefore may be used in conjunction with a dual redundant system. In this form, the dual comparison is used to determine that a fault (or failure) has occurred. The consistency test is then executed to determine which of the data lines is the culprit.

Miscellaneous Advanced Methods

Grouped in this category are a number of fault or failure detection methods what have several attributes in common. They tend to be complex in the required analysis and implementation, they tend to require detailed knowledge of the dynamics of the system being monitored or controlled, they may require a specialized structure of the system (such as linearity with additive, Gaussian, white process and observation noise), and they are generally oriented to detecting failures as opposed to faults. In regard to this latter observation, they are likely to accept invalid data that does not persist and to fail a data line which is suffering from only a momentary sequence of invalid data points. Another factor is that, because of their complexity, they are more appropriate to monitoring a few, or at the very most, a few dozen data lines. In contrast, a ship-borne monitoring and control system could involve hundreds or even thousands of input signals, making the computer burden associated with the advanced methods untenable.

Among the positive attributes of the advanced methods is that many of them are oriented to determining any changes in overall system dynamics. This means that they can be used to detect problems which may occur in the systems being controlled or monitored (i.e., they are not limited to detecting data line faults). Additionally, many of the advanced methods also contain hypothesis testing as an integral part of the procedure, which permits an evaluation of their associated error probabilities.

Because of their complexity and structural constraints, many of the methods may tend to be impractical and of somewhat academic interest, existing only in the literature. To a large extent, they have been supplanted by multiply redundant, self-reconfigurating systems which are more likely to use voting procedures and analytic redundancy than the methods to be described. Therefore, we will present only a concise sampling and refer to Wilsky (Reference 3) for further examples and references. (Note: The following discussion assumes the reader has at least a passing familiarity with estimation theory. However, even if one does not have this familiarity, the general notions should still be evident.)

A well-known property of Kalman filters is that the innovation sequence is Gaussian and white. (The innovation sequence is the difference between the actual observations and the estimated observations and is used as the "input" to the Kalman filter.) If a Kalman filter is used in the monitoring or control system, then the resulting innovation sequence will possess the desired properties unless a failure occurs. Thus, it is possible, in principle, to test the innovation sequence for whiteness and this has been suggested (Reference 4) as a failure detection mechanism. However, in order for this whiteness property to hold, the system being monitored must be linear, the dynamics must be well known (although, again in principle, it is possible to estimate the dynamics), and the process and observation noise must be Gaussian and white. We seldom find such properties in real-life systems, especially in the sea-borne environment. Further, because of the properties of the Kalman filter, a few hard over faults might result in a diagnosis of a failure. They would also corrupt the filter output unless suitably screened by other tests, such as magnitude tests. Another difficulty is that the Kalman filter dynamics introduce a considerable lag in the detection process and will respond quite slowly to even abrupt changes.

Methods to obviate the sluggishness of the Kalman filter approach have been suggested that involve essentially increasing the bandwidth of the filter, thus making it more sensitive to sudden changes in the data (References 5 and 6). However, the drawback of these approaches is that the filters become increasingly sensitive to system noise and are more likely to cause false alarms.

In a similar vein, an estimation method has been suggested for detecting bias failures by including a bias state in the dynamic system and then estimating the bias state vector (Reference 7). Once again, the process is highly complex and is limited in applicability.

One more advanced method will be mentioned before proceeding to the next topic. Several authors (References 8 and 9) have investigated the use of a "bank" of Kalman filters, each constructed based upon a different assumption of a model of the system, with each model corresponding to a particular failure mode. The innovation sequence for each of these filters is then tested to determine the conditional probability that each system model is the correct one. This procedure is limited by the required computational power and constraints on the nature of the underlying models.

HOW MANY FAULTS EQUAL A FAILURE?

Having indicated in general terms how one might determine whether a particular piece of data is faulty, how then can one proceed to determine whether this is a temporary condition which may rectify itself or whether the condition is more or less permanent and a failure should be declared? The answer to this question is not easy and must be approached from the viewpont of the user of the data. How much bad data can the system tolerate before performance is degraded to an unacceptable level? To a great extent, the answer to this question depends on the way in which the invalid data is handled, the robustness of the system, and the criticality of the operation. It is not uncommon, for example, for digital control systems to operate at sampling rates from 2 to 10 times faster than that actually required to maintain good stability and performance. Under the proper circumstances, then, it would be conceptually possible to operate with 50 to 90 percent invalid data. (In practice, a data line that produced this much invalid information would be highly suspect and be subject to corrective maintenance action.)

Simulation is a key tool in evaluating methods for handling the invalid data, and, of relevance to this section, in determining how much good data is required for continued operation. The determination of how much good data is needed is generally done by considering various sequences of invalid and valid data under the most demanding of operational environments. Data sequences in which N out of every M data points are invalid can be tested with both regular, that is, periodic, repetitions of the valid/invalid sequence, as well as with randomly occurring valid/invalids with the probability of invalid data selected to span an expected range (Reference 10).

Additional simulation or analysis will permit a determination of the longest time period without valid data that can be endured without loss of control if a failure is finally declared.

The two criteria established, that is, the fraction of invalid data permitted and the longest period of safe operation without valid data, may then be used to determine when to declare a failure. In a specific example, an automatic control system designed to provide ship steering and propulsion control during underway replenishment operations (Reference 11) was determined to experience noticeable performance degradation if more than one out of every five data samples was invalid (the algorithm was meant for prototype use and was not fully protected). It was also determined that a period of not more than five seconds without valid data could be tolerated. With a sampling rate of eight per second, these two requirements were combined into a single requirement that more than 8 faults from any 40 consecutive data samples would result in declaring the system to be failed.

The N-out-of-M-faults to be declared a failure is a commonly employed method and, for heading control, numbers that have been used under various circumstances are 3 out of 10, 3 out of 8, and 10 out of 32.

A related method to the N-out-of-M-procedure is to use the output of a first order filter to trigger the failure declaration, as follows: the filter input is set to unity (one) if the data is faulted and to 0 (zero) otherwise. If the output of the filter exceeds a predetermined threshold, the data line is declared failed. The threshold and the time constant of the filter can be determined in such a manner as to approximate the N-out-of-M criteria described above, but with substantially less computational burden. This may be important if a large number of signals is being evaluated.

A technique that combines the method just mentioned and the magnitude test for fault detection is to use the data itself as input to a first order filter. On the assumption that the only type of faults that will occur are hard over faults, the output of the filter can be used for failure determination by a proper selection of the time constant and the decision threshold. These parameters, if properly selected, will allow several faults before a failure is declared and are, therefore, related to the N-out-of-M criterion described above. While this procedure has the advantage of simplicity by combining the fault and failure detection in a single algorithm, it is only effective under a limited set of circumstances.

A method that may be used to augment the control related methods described above, and that may have utility in specialized circumstances, is to approach the failure decision problem from a probabilistic point of view. If it is assumed that the data line, once it provides invalid data, is permanently failed, then the problem becomes one of determining whether the decision of a particular fault declaration is correct. If the probabilities of the four possible outcomes of the fault decision process are known (probability of fault detection when no fault exists, probability of no fault detection when a fault has occurred, and their two complements), then it is possible to compute the probabilities of false failure alarm and correct decision based on an evaluation of the past fault decisions. That is, if the probability that a declared fault is actually a fault is high, then the probability that a sequence of N declared faults are false alarms is very low. Specifically, suppose we have established an N-out-of-M failure criteria. If P_{fa} is the probability of a fault false alarm, and P_{vv} is the probability of declaring valid data to be valid, then the probability that N out of M data points will be declared to be faulted when the data line is functioning properly is:

$$P(\text{failure false alarm}) = P_{fa}^{N} * P_{vv}^{(N-M)}$$

Probabilities of other events can be computed in a similar fashion and such computations can be used not only to determine the N out of M criteria, but also to adjust the thresholds of the fault detection procedure, since these determine the fault detection probabilities. That is, in an actual system design, the fault and failure criteria are not treated as independent, but are interrelated through the needs of the system and its users.

WHAT TO DO UNTIL THE DOCTOR COMES

If data have been determined to be invalid and there is no alternate sources of valid data, i.e., the system has no redundancy, then some decision must be made as to what action to take while waiting either for a failure to be declared or for the re-occurrence of valid data. The following possibilities exist and have been used in sea-borne applications.

1. Do nothing. Bypass all control computations until valid data are available and then continue. The advantage of this approach is its simplicity. The substantial disadvantage is that considerable adverse transient behavior might result during the restart process. If the invalid data consist of a single bad data string followed by long periods of valid operations, this may not be a serious problem. However, if invalid data occur randomly and frequently, the consequences can range from mildly annoying to a considerable performance degradation.

2. Continue to cycle all computations using the last good data point. This has the advantage of simplifying the control execution logic and can be effective if the rates of all controlled variables were small

at the time the fault occurs. However, if rates were significant, continued execution with the last good data point can lead to large errors in control. Errors can also be aggrevated if the last "good" data actually contained substantial noise that did not exceed the detection threshold.

3. Continue processing with the invalid data. The premise that has resulted in the implementation of this approach is that faults would be rare events and their effects would be mitigated by filters in the control law. The other side of the filter attenuation coin is that the filter permits the influence of the invalid data point to persist for a period of time, depending upon the filter characteristics.

4. Use predicted data. If the fault detection scheme employed a data extrapolation method, then it is a natural procedure to use the predicted value in place of the invalid datum. The length of time for which the prediction is valid and the basic accuracy of the prediction depend on the quality of the dynamic model used in the diction process. A potential hazard associated with using predicted data is that data quality might not be good during the period just prior to the detection of the fault due to gradual deterioration. If this is true, then the prediction might be extremely poor.

5. Estimate the data using alternate information sources. If the data are analytically redundant with other data, then they may be estimated computationally. This is a natural extension if data consistency tests are employed for fault detection as the computational procedure will be in place. The primary disadvantage of this is exactly the same as that pointed out with regard to consistency tests. That is, they may impose a substantial computational burden. However, this may be avoided by only executing the estimation procedure when required.

6. Adjust the control algorithm coefficients to account for the increased time step effectively introduced when data are not available. A simple example will clarify this concept. If a derivative is estimated by dividing the difference between successive samples by the time step, then if a data point is missing, the effective time step is increased and the derivative is estimated by dividing successive valid data by twice the time step. In most real systems, the actual procedure will be more complex, as filter coefficients are frequently exponential functions of the time step and this exponential function will have to be recomputed as required. This may not prove to be a problem if filter coefficients are computed every time step for other reasons, such as in speed adaptive filters. Additional problems may arise if filters of second order or higher are used, as the past values must be equally spaced in time. In most cases, these higher order filters can be replaced by N first order filters, thereby avoiding the unequally spaced data point difficulty.

Despite the apparent complexity of this approach, it has been used successfully in applications where the filter coefficients were computed at each time step so that the inclusion of a variable time step size was not difficult. The major disadvantage of this method is that not all filters and control procedures are amendable to varying time step sizes, even when such variations occur only rarely.

SUMMARY

The increasing complexity of ship-borne control and monitoring systems has led to a situation in which tremendous amounts of data are transmitted and processed. In most cases, it is vital that the control system continue to function in the presence of temporary or long-term malfunctions that occur in the control and monitoring system. This paper has focused primarily on one element in the development of fault-tolerant system operation. That is, the determination of when a data line fault has occurred. This is only the first step in what may be a highly sophisticated process that will serve to isolate the failure, reconfigure the system to operate around the failed member, and notify maintenance personnel of required corrective actions.

The purpose of the paper is not to advocate any particular approach, nor even to claim that it represents a comprehensive survey of available techniques and methodology. This would require more time and space than is available. The primary goal of the paper is to call attention to a problem that often receives only cursory consideration until difficulties occur at sea. The secondary goal is to present several of the methods that have been used or suggested and provide some insight on their applicability based on experience with several fault/failure detection systems.

REFERENCES

1. Deckert, J. C.; Desai, M. N.; Deyst, J. J.; and Wilsky, A. J. F8-DBFW Sensor Failure Identification Using Analytic Redundancy. IEEE Trans. on Auto. Cont., Vol. AC-22, October, 19774.

2. Broen, R. B. A Nonlinear Voter-Estimator for Redundant Systems. Proc. 1974 IEEE Conf. on Decision and Control.

3. Wilsky, A. S. A Survey of Design Methods for Failure Detection in Dynamic Systems. Automatica, Vol. 12, 1976.

4. Mehra, R. K., and Peschon, J. An Innovations Approach to Fault Detection and Diagnosis in Dynamic Systems. Automatica, Vol. 7, 1971.

5. Fagin, S. L. Recursive Linear Regression Theory, Optimal Filter Theory, and Error Analyses of Optimal Systems. IEEE Int. Conv. Record, March, 1964.

6. Tarn, T. J., and Zaborszky, J. A Practical Nondiverging Filter. AIAA Journ., Vol. 8, 1970.

7. Friedland, B., and Grabousky, S. M. Estimating Sudden Changes in Biases in Linear Dynamic Systems. IEEE Trans. on Auto Cont., Vol. AC-27, February, 1982.

8. Athans, M., and Willner, D. A Practical Scheme for Adaptive Aircraft Flight Control Systems. Symposium on Parameter Estimation Techniques and Applications in Aircraft Flight Testing, NASA Flight Research Center, Edward Air Force Base, April, 1973.

9. Lainiotis, D. G. Joint Detection, Estimation and System Identification. Inform. Control, Vol. 19, April, 1971.

10. Chan, Y. K., and Edsinger, R. W. A Correlated Random Numbers Generator and Its Use to Estimate False Alarm Rates of Airplane Sensor Failure Detection Algorithms. IEEE Trans. on Auto. Cont., Vol. AC-26, April, 1981.

11. Ware, J. R., and Best, J. F. Control System Design for the Close Range Ship Control System, Vol. II: Control Design Procedures. ORI Technical Report 1991, April, 1982.

ADDITIONAL BIBLIOGRAPHY

Athans, M.; Whiting, R. H.; and Gruber, M. A Suboptimal Estimation Algorithm with Probabilistic Editing for False Measurements with Application to Target Tracking with Wake Phenomena. IEEE Trans. on Auto. Cont., Vol. AC-22, June, 1977.

Kerr, T. H. Statistical Analysis of a Two-Ellipsoid Overlap Test for Real Time Failure Detection. IEEE Trans. on Auto. Cont., Vol. AC-25, August, 1980.

Hibey, J. L. Performance Bounds for Differential Systems Incurring Abrupt Random Changes. IEEE Trans. on Auto. Cont., Vol. AC-26, June 1981.

Ware, J. R. Rejection of Invalid Data for an Integrated Monitoring and Control System. ORI TECH MEM 112-81, March, 1981.

Ware, J. R. Memorandum on Data Prediction and Validation Algorithms for TRIDENT NAVDATA, ORI TECH NOTE 83-1, August, 1981.

Ware, J. R. Comments on Trend Analysis Procedures. ORI TECH MEM 114-81, March, 1981.

Failure safe design and maintainability of modern automation

by Ir. J.A.M. ter Horst

Naval Electronics and Optics Maintenance Establishment
Royal Netherlands Navy

CONTENTS

1. INTRODUCTION

With help of modern technology, more and more routine functions are automated, while other functions are realized in a more optimal, but also more complex, way.

It is not unusual, that almost all design effort for those applications is concentrated on the normal operation of a system. The effects of degraded or faulty operation of the system under control, conventionally taken care of by the flexable response of a human controller, are mostly insufficiently attended. Also the effects of failures in the complex control equipment are not deeply investigated.

Without any automatic alarm facility, failures in the system under control or the controller itself can only be detected by a human supervisor, as a result of gross performance errors or any secundary effect resulting from such a failure. Therefore most controllers have some form of automatic alarm, being activated when the system performance degrades below a certain threshold. In that case a human operator can, in case the failure is in the controller itself, disable this controller and start manual control over the system. If the failure is in the system under control, the operator can undertake adequate action to prevent catastrophic results.

For instance an automatic pilot of a ship is normally equipped with an off-course alarm, that sounds a bell when the difference between actual and setpoint course exceeds a certain threshold. Such an approach causes the following disadvantages :

1) A relative high operator attendance is required to avoid catastrophic results of occasional failures.

2) When the alarm is activated, there is a fair chance that the operator will find the system with the output of the controller in some extreme condition, causing a rapidly increasing system error.

3) In case of a defective automatic controller, a trained operator shall be available on short notice to perform adequate manual control. A need for trained operators requires regular training on the system itself or on a special trainer. As an alternative a degraded system performance in the manual control mode can be considered.

4) It is not always guaranteed that failures in the external powersupply or a sensor for the actual controlled parameter will cause an alarm.

To avoid the last mentioned problem, the alarm shall be based on independant or redundant external power and feedback of the primary actual system parameter(s).

The disadvantage mentioned under 3) can be eliminated by a redundant set up of the control system, including the external powersupply and sensor configuration.

The disadvantage mentioned under 2) can be significantly reduced by the implementation of a Performance Monitor, capable of detecting a failure in a very early stage and causing an automatic disable of the controller, with the controller-output in a neutral or least dangerous condition.

All disadvantages mentioned above can be eliminated using a full redundant control system with a Performance Monitor effecting, in case of a failure in the active subsystem, an automatic switch over to a hot standby subsystem.

The conventional way to increase the reliability of systems has been the use of high quality (MIL SPEC) components, together with extensive inspection and burn-in programs. Compared to this approach a dual redundant system, with automatic reconfiguration and the use of good standard industrial components and production techniques, will show significant higher figures for functional Mean Time Between Failures (MTBF). This functional MTBF relates to the conditional probability, that a new failure will occur when the other redundant subsystem has become defective already and has not yet been repaired. Besides this very low probability of a functional failure, this failure will always be preceded by a warning, being the first failure causing the system to loose its redundancy.

Unlike in airborne or space environments, ship systems can be repaired on board, restoring the full redundant availability after a failure has occurred. Because this approach requires twice the amount of hardware with standard reliability level, the equipment MTBF, related to the mean time between events requiring attendance of a technician, will be lower than that of a singular system concept of extended reliabilty. The Mean Time To Repair (MTTR) can be kept to a minimum by the use of adequate Built In Test (BIT) for Fault Location (FL) purposes. This will help the full redundant system availability to be optimized and the remaining probability of functional failures, as a result of simultaneous presence of more than one equipment failure, to be minimized.

This paper deals with an approach, to achieve failure safe systems and their maintainability. After an investigation into the various failures that can be expected and an introduction in the various possible techniques, that can be used for automatic detection of failures, examples of implementation Performance Monitoring (PM) functions and repairability into a system are given.

2. EVALUATION OF FAILURES TO BE EXPECTED

The first step to obtain failure safe systems is a detailed analysis of all expectable failures, in each part of the total system, and their effect on system performance. This analysis shall be based on experience in the failure oehaveour of similar or comparable systems, subsystems, modules, components or technologies. The amount of effort involved depends highly on the quantity and quality of already available experience and how closely this relates to systemlevél for the Performance Monitor aspects and replacable module level for the Failure Location aspects. In worst case all expectable failures of individual components and other parts, like wiring and connectors, have to be evaluated on effects on module level and further on effects on system performance level. All relevant failures can be classified in the following categories :

2.1 Suddenly Occurring Hard Failures

This type of failures cause a steady and significant reduction of system performance or, if applicable, the full redundant availability of the system. This type of failures require, at least for Performance Monitoring purposes, full coverage by on-line BIT facilities.

2.2 Non Critical Hard Failures

These are failures that not cause significant reduction of system performance or redundancy. Where it is not practically, to cover those failures by on-line BIT facilities off-line go/no-go type periodic operability tests can be considered. As an example a display and lamp test can be mentioned.

2.3 Slow degradation of system performance

Slowly decreasing system performance, as a result of wear and tear or slow changes in component parameters, will normally not be covered by BIT facilities, but by scheduled maintenance procedures. The accuracy requirements for BIT facilities being able to detect those performance degradations would be dramatically higher than those, required for detection of suddenly occurring significant reduction of system performance.

2.4 Incidental Short Lasting Failures

Examples are power interruptions, or failures in analog circuits, lasting less than e.g. 300 m.sec , or occasional parity or framing errors in serial digital interfaces. This type of failures shall not cause long standing degradation of system performance, or operator alarms. It may be helpfull to count this type of failures in a confidence counter, to be used by the system maintainer.

2.5 Spurious Failures

These failures, also called "soft" or "weak" failures, can be caused by for instance unexpected Electro Magnetic Interference (EMI) effects, or remaining hidden software errors. Because these failures are not expected, they will not explicitely taken care of by normal BIT facilities. Furthermore they can be very difficult to diagnose, because they happen only very rare and,in general, only during normal system operational conditions. Therefore a system shall have facilities for connection of special computer and/or analog test equipment in such a way, that this has not a significant impact on normal operational use of the equipment.

3. AUTOMATIC DETECTION OF FAILURES

3.1 General Design Possibilities

Various strategies can be used for automatic on-line detection of failures in various well defined parts of a system. Which solution is to be selected, for each detection requirement, depends fully on the particular circumstances of the system-part to be monitored. So an optimal solution, providing adequate confidence at a minimum cost, shall be made on a case by case basis.

Several detection aids will also generate an alarm when they fail themselves. However, especially in tail-ends of failure detection circuits, failures may occur without notice. Instead of unreasonable BIT on BIT complexity, periodic simple go/no-go check procedures are suggested.

Possible aids for automatic failure detection in general are :

3.1.1 Watch Dog , also called "hartbeat", monitor. This is a retriggerable monostable multivibrator function, that monitors the periodic occurance of events. Such a monitor will be used to verify the periodic excecution of a controller routine, or the presence of a clock or carrier frequency.

3.1.2 Window Comparator , a circuit that is able to verify an analog signal to be, within certain tolerances, equal to a fixed or variable reference signal .
An example of such a comparator is given in figure 1 , using half of a quad orable op-amp integrated circuit. The BIT TESTBUS is a centralized aid for go/no-go testing of a number of these or similar BIT circuits.

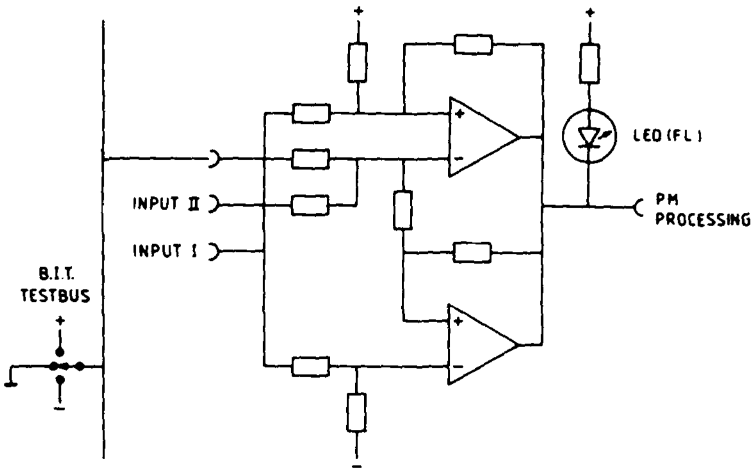


Figure 1 WINDOW COMPARATOR

3.1.3 Forward Duplication , where a function to be monitored is mechanized twice, and the output of both channels is continuously compared. It should be noticed that this is not a redundant set-up, because only one channel is always used in the system, while the other channel is always used for monitoring purposes only. In general this last channel can also be less complex, because its accuracy requirements are lower. This because the function of BIT is normally not a continuous verification of accuracy, but a detection capability for serious failures, probably causing serious functional system performance degradation. So, for instance, a high precision AC to DC signal converter can be monitored by a simple peak rectifier. In figure 2 the principal of this BIT technique is illustrated.

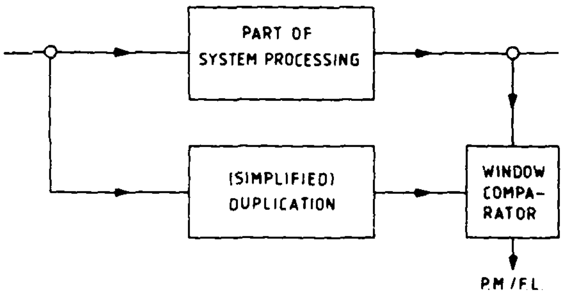


Figure 2 FORWARD DUPLICATION

3.1.4 Inverse Duplication , using the inverse mechanization of the function to be monitored, in a feedback monitor channel. This principle is illustrated in figure 3 .

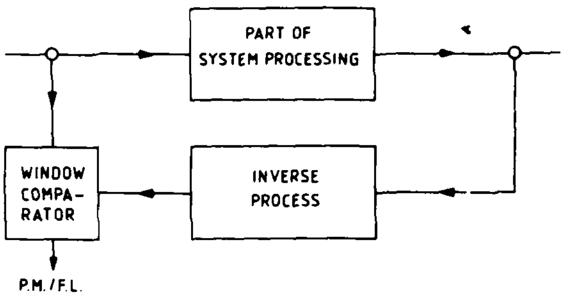


Figure 3 INVERSE DUPLICATION

3.1.5 Hybrid Duplication , being a mixture of the two preceeding possibilities.

3.1.6 Use of Testsignals , being processed, together with operational signals, by a systempart to be monitored. To eliminate interference between operational and test signals, the following separation techniques can be used :
-) frequency division multiplex (also usable for "dither" excitation, if required)
-) time division multiplex (e.g. in digital computers)
-) the use of pseudo random noise as testsignal and correlation techniques for detection
-) compensated testsignals, as illustrated in figure 4

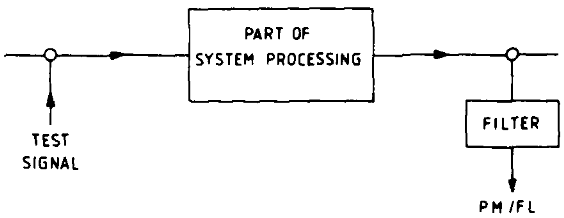


Figure 4 USE OF TESTSIGNALS

Except in cases where time division multiplex is involved, the channel being monitored shall not show non linear transfer functions, like e.g. saturation as a result of short lasting excessive error signals in a servo loop.

3.1.7 Secundary Parameter Verification , like :
-) correct carrier signature (see also para 3.5.3)
-) parity or CRC checks
-) framing errors in serial digital information

3.1.8 Reasonableness Tests , like e.g. verification of the maximum reasonable rate of change of certain parameters.

3.2 Availability of Power

Monitoring of external power inputs will help failure diagnosis, in case of external power failures. In case of three phase electrical power, each phase shall be monitored separately. Also redundant power configurations shall be monitored independantly, because drop out of one of those may have no effect on normal system operation and will remain hidden untill a second failure will cause a serious problem. A simple example of such a monitor mechanization is given in figure 5 .

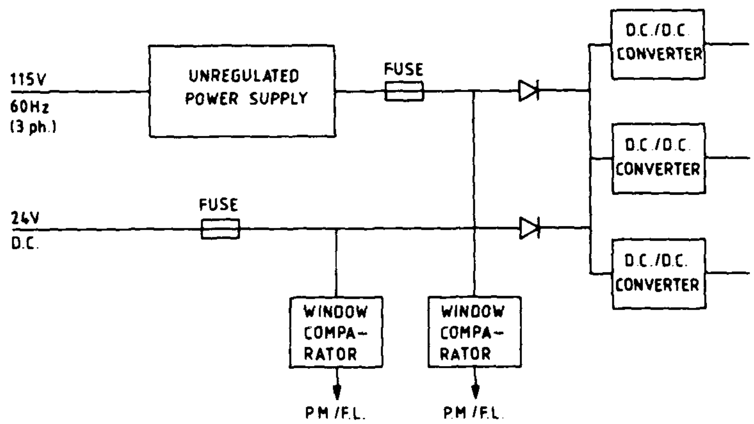


Figure 5 EXTERNAL POWER MONITOR

Special attention may be required to distinguish between short power interruptions, as e.g. related to switch over operations in the external power network, and longer black out conditions. It is likely that there are different requirements for the status, in which the system is initialized, after short or long power interruptions.

Also internal power converting equipment should be monitored for Failure Location purposes.

3.3 Digital Computer Equipment

The normal task of a computer in a control system is the periodic excecution of an operational program. This program can be composed of various loops, running with different repetition frequencies. Furthermore this program will probably include correct handling of spontaneous external events, like changes in mode of control, setpoint or alarmstatus of equipment outside the computer. In the initialization after start-up and after the detection of a computerfailure, the computer can excecute off-line test and failure localizing routines. It is not unlikely that a computer is used as a core of a total Performance Monitor and Failure Localization concept. Especially in those cases, this computer shall be equipped with rigid self monitoring capabilities. These capabilities should include :

3.3.1 Watch Dog Concept . As soon as a computer has entered a normal operational condition, a watchdog (hartbeat) concept shall monitor whether the computer actually excecutes the various cyclic program loops. Where critical situations can result from a computer not responding to interrupts caused by external events, these interrupts should also start a time-out function that monitors a handshake signal, to be generated by the computer. This time-out function and at least the tail end of the watchdog concept shall be realized separately from the computing function. Operator inputs, like changes in setpoints or operational mode, can be monitored by this operator.

3.3.2 On-Line Checks can monitor, whether the results of the computations can be considered to be free of errors. The following aids may be helpfull :
-) extensive internal parity verification
-) reasonableness checks
-) on-line testroutines
-) loop around checks on output and (multiplexed) input channels

3.3.3 Off-Line Test and Diagnostic Routines . In case of a computer failure, not being limited to one repetitive program cycle only, a set of off-line testroutines, to verify correct CPU, RAM and PROM operation, can be started. This type of tests are also excecuted during normal start-up procedures. In case of (redundant sub-)system failures, detailed I/O diagnostic routines can be started to verify correct operation of complete I/O functions, so also including all computer external hardware.

3.3.4 Built-In Hardware Diagnostics . A detailed analysis of integrated testability of computer hardware goes beyond the scope of this paper. Reference 1 contains a special report, dealing with this subject specifically.

3.4 Analog Circuits

The performance of analog parts of the system can be monitored by aids, as mentioned under 3.1 . Compared to a highly computer centralized set-up for detection of failures, the use of distributed local detectors, based on analog circuits, may reduce over-all system complexity. Also, when certain analog circuits remain active in a computer down emergency condition, these local detectors remain functional.

3.5 Sensor Inputs

Failures in external sensors can be very critical, because they may lead to serious misinterpretation of the actual status of the total system. Some sensor monitoring techniques are :

3.5.1 Sensor Duplication . Provided that no common failure sources, like e.g. a common link between the system parameter to be measured and the two sensors or disastrous environmental conditions, can be relevant, the probability of two sensors showing identical errors simultaneously can be neglected. Special attendance should be given to the power concept of such an arrangement.

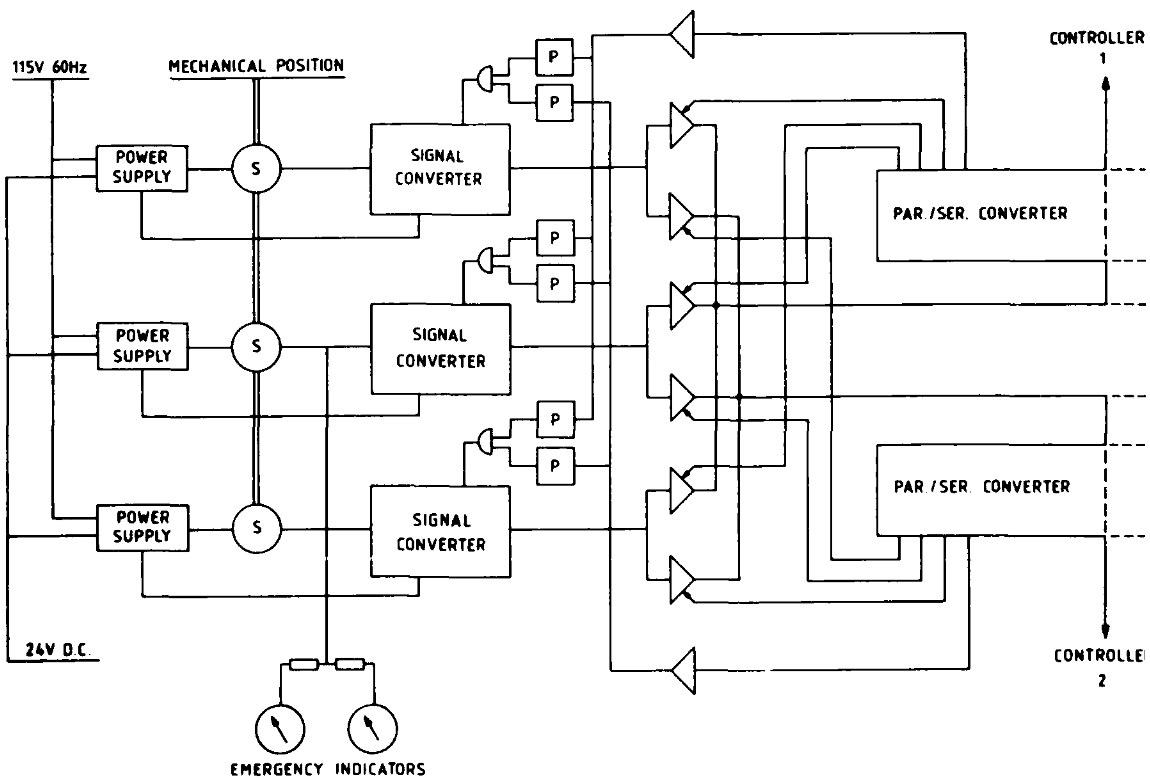


Figure 6 SENSOR SUBSYSTEM

Furthermore the final comparison between the two sensors should be done in the computer, being on-line monitored itself, in order to be sure that also the signal conversion and computer interface part operate correct. It should be mentioned, that in a dual redundant controller set-up, this solution implies the use of a total of four sensors. An alternative for that configuration is the use of three independent sensors with a majority vote mechanism in each of the two controllers. This can also be mechanized in a redundant sensor subsystem, as shown in figure 6 . The small blocks marked "P" take care of a short freeze of data, to enable reliable transfer to the par/ser converter. Overlaps of freeze periods, initiated by the two par/ser converters do not cause any problem. However, two successive freeze commands from anyone of the par/ser converters individually shall include a dead time interval, in order to eliminate any possibility for transfer of stale data to the other converter. The right hand side of the figure indicates, that more data can be combined into one serial digital datalink to a controller. This link itself shall be monitored e.g. by time-out and parity or CRC checks. The probability that a serious system problem, causing a need to use emergency indicators and control, coincides with a non availability of a particular singel sensor and/or those indicators, can be neglected. Therefore the remote emergency indication on two locations is mechanized in a non redundant way.

3.5.2 External Validity Information can be derived from complex sensors with built in Performance Monitor. An example of such a sensor is a horizon stabilized compass, having enough built in checking capability to certify its heading output to be true or false. It should be clear, that this validity information does not cover the transfer of data from the sensor to the controller. So that part remains to be monitored by a separate detector. See also figure 7 .

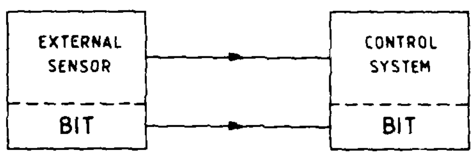


Figure 7 EXTERNAL VALIDITY INFORMATION

3.5.3 Verification of Secundary Parameters . In frequency modulated transfer of information, the carrier amplitude is a secundary parameter, that can be monitored to verify the confidence level of the information link. Another example of verification on a secundary parameter is given in the left part of figure 8 . Here a remote synchro sensor is checked on correct relation between reference voltage and a specific combination of S-line voltages. It can be shown that an electrical failure, like an open connection or some form of short circuit, that influences the relation the S-line voltages, being the actual carrier of the angular information, will also change the square sum of the two sinewave amplitudes at the output of the SCOTT-T transformer. In other words, when the direction of the angular vector is affected, also the absolute size of this vector, in relation to the reference excitation, will be affected. This effect is used to provide a reliable check on the performance of this remote sensor.

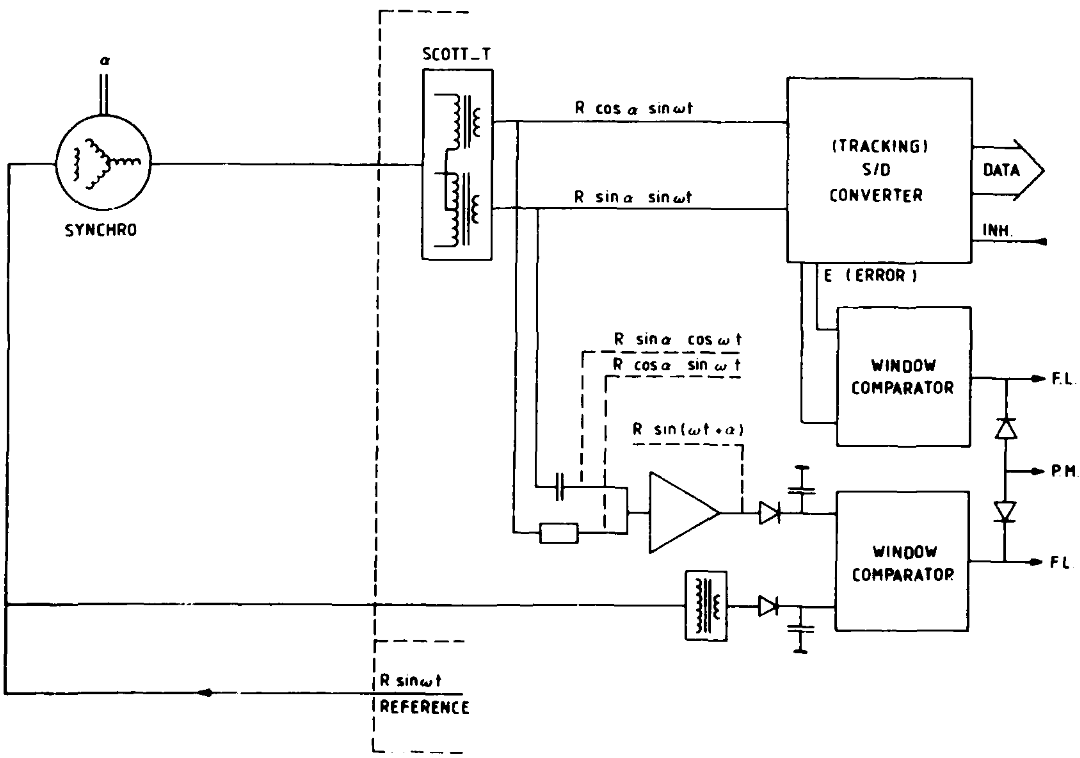


Figure 8 MONITOR ON SYNCHROSIGNAL . TRACKING CONVERTER ERROR

3.5.4 Impedance Checks can be used to detect open or short circuit conditions in remote sensors, or their connecting cables. This type of monitoring is further highlighted under para 3.6 .

3.5.5 Test Signals , as indicated under 3.1.6 can be used in some applications, like e.g. a temperature sensor in a relative stable environment. Using a local small heater in the direct vicinity of the sensor, a periodic short thermal pulse from that heater shall also be detectable in the controller. The detection criterium will be based on the short presence of a normally impossible rate of change. The effect of the pulse on the controller process can be eliminated by counteracting compensation. See also figure 9 .

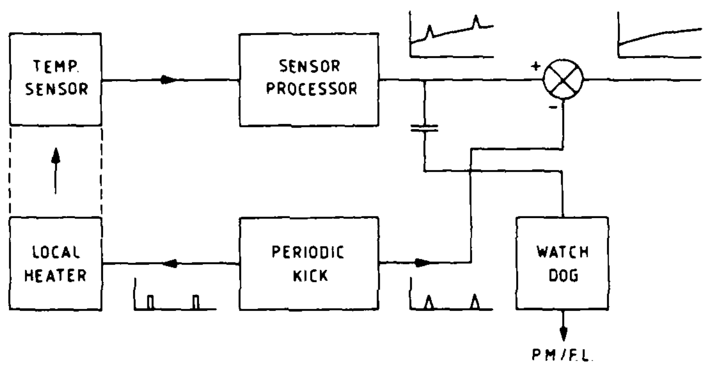


Figure 9 COMPENSATED TEST SIGNAL

3.5.6 Reasonableness Checks can provide detection of failures to a reasonable confidence level. This type of monitor will normally be based on surpassing of a maximum expectable rate of change and/or a correct trend. Hard failures often result in a step in the related signal or a steady signal, where it is evident that a change should occur.

3.6 Control Outputs

Besides verification of a control signal as it leaves the controller, it can be impor tant to check whether this signal has been received in correct form by e.g. an actuator. This information can be used to avoid failure diagnosis to be started at the wron location by the wrong technician. Using, as an example, a solenoid control of an electro/hydraulic actuator, the following possibilities can be discussed :

3.6.1 Impedance verification , checking a correct current to voltage relation, using the normal operational (DC) control signal. Any open or (partially) short circuit, that may seriously affect normal operational performance, will be alarmed. Figure 10 illustrates this monitor option.

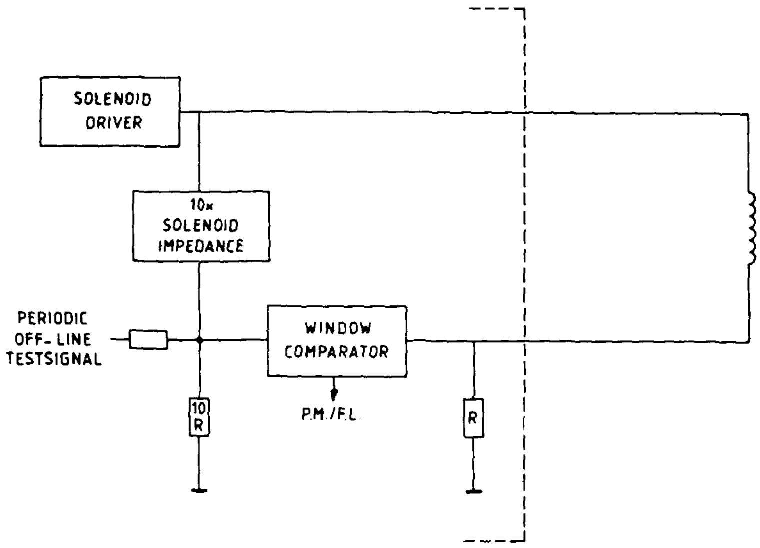


Figure 10 MONITOR ON EXTERNAL IMPEDANCE

3.6.2 Impedance Verification on Test Signal . As indicated in para 3.1.6 , special testsignals can be used in addition to the normal operational signal. Compared to the solution given under para 3.6.1 , the following advantages of this approach may be relevant :

-) the monitor function also operates in a hot standby situation, where no actual excitation of actuators is allowed
-) the testsignal can also be effective as dither signal, usable to avoid stiction
-) some solenoïds show significant changes in AC impedance, as a function of the core position, which can be used to check also the position of the core (feedback function as indicated under 3.1.4)

The off-line testsignal, as shown in figure 11 to be used for periodic operability verification, should be derived from the same source as the testsignal itself.

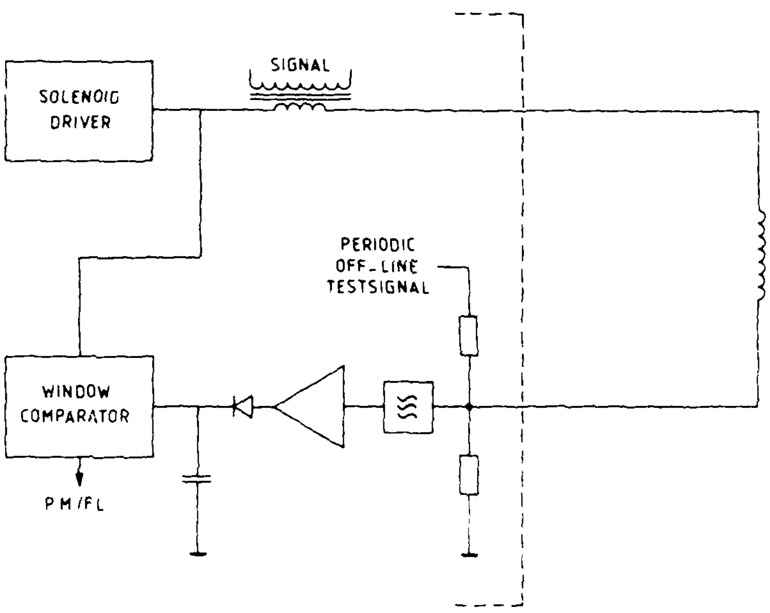


Figure 11 SOLENOID IN/OUT IMPEDANCE MONITOR

4. PERFORMANCE MONITOR

A Performance Monitor (PM) is that part of an on line Built In Test configuration, that provides, to an operator or watch of a system, the following information :

-) an alarm or warning of adequate urgency level in case of any failure in the total system, encompassing the automatic (redundant) controller together with all external sensors, actuators and power sources and the system under control
-) a clear indication about the remaining total system availability status
-) in case of complex and spatially dislocated systems, a clear indication about the location of the failure and the type of technology being involved, in order to be be able to send the right technician to the right location to start further diagnosis and repair
-) a signal that permits, in case of a failure that may significantly degrade system performance, automatic reconfiguration of a redundant system (fail operational concept), or automatic switch over to a least catastrophic interim condition, from where manual control can be started (fail safe concept). Simple comparison between comparable signals in a dual redundant system configuration can never be used, to detect which of the two subsystems is defective. However, a mojority vote concept in a triple redundant configuration can perform this function. In some cases a majority vote concept, implemented in both subsystems of a dual redundant system and based on comparison of comparable systemparameters between those subsystems and a reference model, implemented in each subsystem, may provide adequate confidence levels for automatic reconfiguration.

False alarms, as a result of non significant and/or short lasting failures or occasional bit errors, should be eliminated as much as possible.

With reference to the point about clear indication about remaining system status, the American Airlines DC-10 accident at Chicago in 1979 should be mentioned. This aircraft could have been saved, after separation of the left wing engine and loss of control over the left wing outboard slats, when the Performance Monitor of the aircraft would have continued to provide information about slat status and stall warning regarding the left wing of the aircraft. This information could have been restored by pilot action, but was not anticipated of prime priority in the hectic 25 seconds between engine separation and final stall of the left wing of the aircraft. This misinterpretation resulted from the fact, that the pilot did not and could not know, that the left wing engine had separated instead of having an ordinary black out, and so concentrated on other alarmindications. As a result of this accident, now at least the slatposition and stallwarning sensor subsystems have to be fully redundant. See also Reference 2 .

To achieve a Performance Monitor of adequate confidence level, a Performance Monitor Plan shall be based on the system functional diagram. In figure 12 an example of such a diagram is given for a ship's steering system. For each significant mode of operation, this plan should result in a table, listing in rows of the first column all significant failures in each of the functional blocks, who may result in serious degradation of system performance. The second column lists the probability of that failure in events per 10^9 hours, eventually multiplied by a seriousness factor being one for serious failures down to zero for irrelevant failures. Further columns are reserved for individual failure detection mechanisms. Each intersection of the resulting matrix indicates the probability (normalized on one) that the particular failure will not be detected in time by the related detector. So non relevant relations, having no detection capability at all, will score "1" , but are left open in the matrix. In case a detector itself is monitored by a periodic operability test, this probability should include the probability of a defective detector at the moment of the related system failure. On the right-hand side of the table some columns are used to list, for various combinations of detectors, the final probability, that a particular failure may occur without being detected. This is normally not a simple linear calculation, because the non detection probabilities of the various detectors can be positively or negatively correlated. The vertical sum of such a column shows, for the related combination of detectors, the remaining probability of a non-detected system failure, potentially multiplied

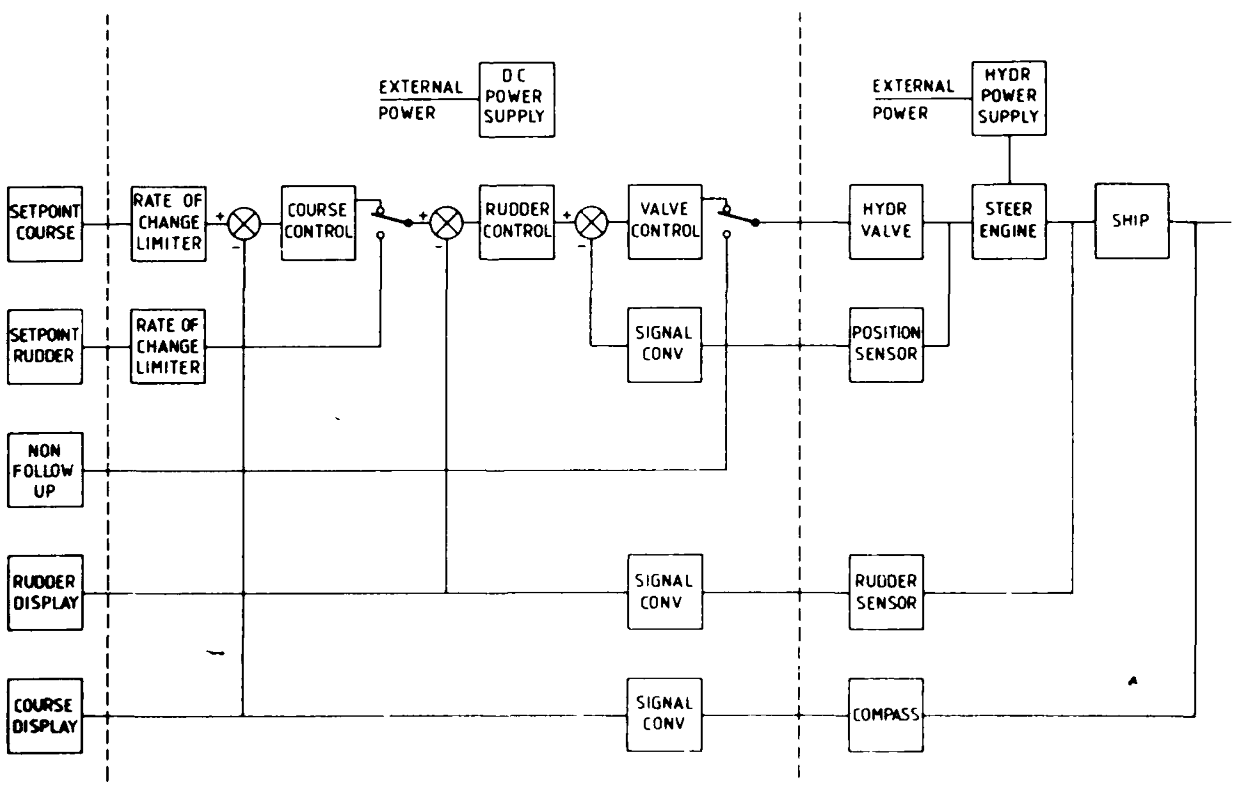


Figure 12 SYSTEM FUNCTIONAL BLOCK DIAGRAM

Failure detector on / Failure source	failure probability in 10^{9} Hr	D.C. power	hydraulic power	course setpoint + rate filter	course error	course controller	rudderangle error	rudderangle controller	valve position error	valve position feedback	rudderangle feedback	course feedback	1 + 4 + 11 IMCO	1 + 5 + 7	all detectors except 4
electrical power	10^{4}	10^{-6}											10^{-2}	10^{-2}	10^{-2}
hydraulic power	10^{4}		10^{-6}										10^{4}	10^{4}	10^{-2}
setpoint course	10^{4}			10^{-6}									10^{4}	10^{4}	10^{-2}
courserate limiter	10^{4}			10^{-6}									10^{4}	10^{4}	10^{-2}
course controller	10^{5}					10^{-6}	.5						10^{5}	10^{-1}	10^{-1}
rudderangle controller	10^{5}							10^{-6}	.5				10^{5}	10^{-1}	10^{-1}
hydr. valve controller	10^{4}								10^{6}				10^{4}	10^{4}	10^{-2}
hydraulic valve	10^{5}								10^{6}				10^{5}	10^{-1}	10^{-1}
valve pos. feedback	10^{4}								.5	10^{-6}			10^{4}	10^{4}	10^{-1}
steering engine	10^{4}												10^{4}	10^{4}	10^{4}
rudder feedback	10^{4}						.5				10^{-6}		10^{4}	10^{4}	10^{-2}
course feedback	10^{4}											10^{5}	10^{-1}	10^{4}	10^{-1}

Figure 13 PERFORMANCE MONITOR PLAN

by a seriousnessfactor. Figure 13 shows an example, using fictive probability figures, of such an approach for the automatic mode of the steering system 12 . The detection criterium in this example is the detection of a failure, before it can cause a course error or a change in rudderangle of greater than 2 degrees. The use of a good model of the rudderangle control loop may reduce the steering engine undetected failure probability by the rudder error detector to 0.1. This will reduce the undetected system failure probability in the last column by a factor of 10 .

This picture will change significantly, when changes in rudderangle up to 5 degrees are accepted, because in that case failures in the valvecontrol loop , the rudder controller and the steering engine are taken care of by the rudder error detector. Without limitation on rudderangle behaviour and maximum allowable course errors in the order of 5 degrees, the IMCO required combination of detectors 1 . 4 and 11 , will result in an undetected system failure probability of 2.10^4 in 10^9 hours. This is primarily caused by the setpoint course and courserate limiter functions. Addition of detector 3 to this configuration will reduce that probability to the order of one undetected failure in 10^9 hours.

A comparable verification should be set up for the manual follow-up rudder control. The non-follow-up emergency rudder control can only be checked by periodic operability verification.

This approach for verification of the confidence level of a Performance Monitor shows, that weak points can be detected and the effect of various detector combinations can be demonstrated. Also it is clearly shown, that the confidence level of certain detectors can be reduced to some degree without noticable impact on the undetected failure probability of the total system. This can result in relative low equency periodic operability checks for those detectors, that require such an approach to verify their function.

5. FAIL OPERATIONAL CONFIGURATIONS

In fail operational configurations the occurrance of a single failure, also when this this is followed by a sequence of other failure conditions induced by this first one will result in a system alarm, but not in a significant reduction of system operational performance. This results in lower alertness requirements for human watchkeeping and no urgent requirement for a trained operator, being capable to perform manual control in case of a controller failure. However, in case of a failure in a dual redundant system, there is no longer a fail operational redundant system available untill that failure has been repaired. In that period of time the system is still fail safe, because the occurrance of a second failure will be alarmed and a least dangerous interim condition can be effected. However, critical manoeuvres, like e.g. refuelling at sea operations, may not be started untill the failure has been repaired. Also the alertness of human watchkeeping should be increased in such a period of time. It is clear that a short Mean Time To Repair is essential to minimize the duration of those conditions.

Critical areas in fail operational configurations are interchange points, where redundant power or sensor information is made available to each of the redundant controller sections. In those situations it should be secured, that no singular failure can seriously degrade system performance. Another critical area is, where redundant subsystems come together to control a singular actuator or display.
Possible solutions for this last problem area are :

5.1 Full Additive Configuration

In this configuration a number of identical controllers, with independant sensors, actuators and external powersupply, operate fully independent and the actuator outputs are added in the system under control. A requirement for this configuration is, that one controller can go defective, causing either maximum or minimum output to the system under control, without degrading systemperformance.

The Performance Monitor of such a configuration is based on comparisons between the various controller outputs. Because this function is completely independent of the controller configuration, the related hardware can have a different redundancy level.

Figure 14 shows, as an example of this approach, a temperature controlled environment. Similar examples can be given in level or pressure controlled systems. Also multi computer, multi task oriented systems use basically this additive approach.

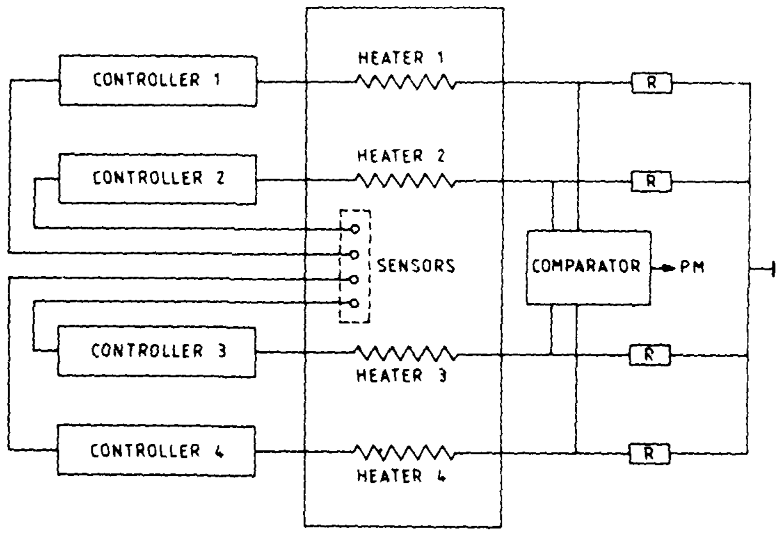


Figure 14 ADDITIVE REDUNDANCY

5.2 Additive Configuration With Exclusion

When the possible saturated output of a defective controller is not acceptable, a Performance Monitor can disable the output of that controller. Figure 15 shows the principal block diagram of such a configuration, as used in BOEING 747 aircraft for CAT III automatic landing with only 50 meter visability. All three autopilots are normally operational, but each one provides only one third of the required force to actuate a plane. When the Performance Monitor detects one force measurement not being consistent with both other ones (majority vote principal), the related output is disabled and the other two autopilots will increase their planecontrol gain factors.

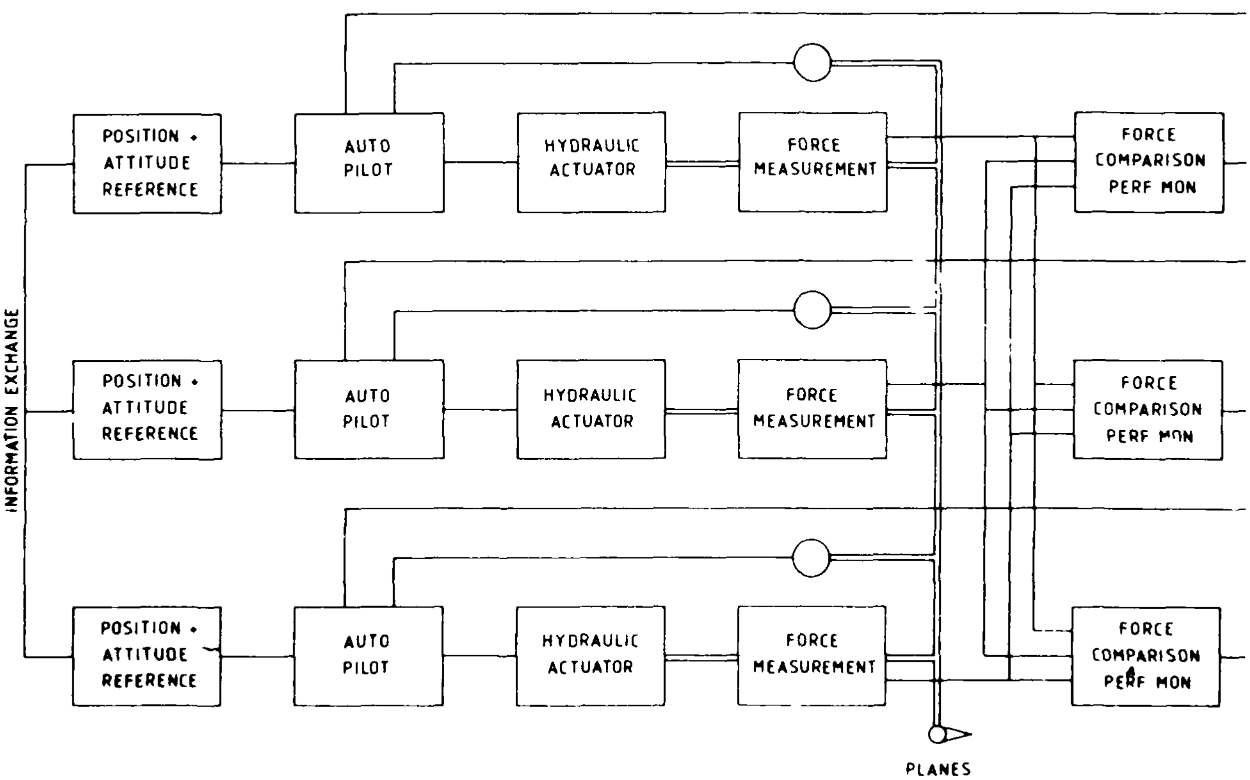


Figure 15 BOEING 747 AUTO PILOT CONFIGURATION

5.3 Switch Over Configurations

When an additive configuration is impossible, a switch over configuration has to be implemented. This implies the use of "active" and "hot standby" subsystems plus a switch-over mechanization. A disadvantage of this solution is the extra dificulty, to verify the correct functional ability, with adequate confidence level, of the "hot standby" subsystem. Another disadvantage is the fact, that a switch-over machanism always adds some non redundant failure sources to the system.

The switch-over function can be divided in three subfunctions :
-) the actual (relais or solid state) switches
-) a memory, that should be non-redundant to avoid misalignment conditions
-) a memory control, being based on a failure detection in the active subsystem, a majority vote by two hot-standby subsystems, an independent monitor or an operator command.

The memory function should be singular, but can be transferred always to the active subsystem, disabling the hot-standby subsystem as long as no failure has been detected in the active subsystem.

Figure 16 shows a symmetric switch-over configuration, using a polar relais. Whether the memory function is implemented in the polar relais or in the active controller, depends on the type of currentcontrol of the excitation coils being pulstype or continuous DC . The actuator outputs of both controllers transmit a regular control signal plus a, controller specific, testsignal (see also para 3.1.6). In that way both controllers can detect the actual position of the relais contact, that actually conducts the operational control signal.

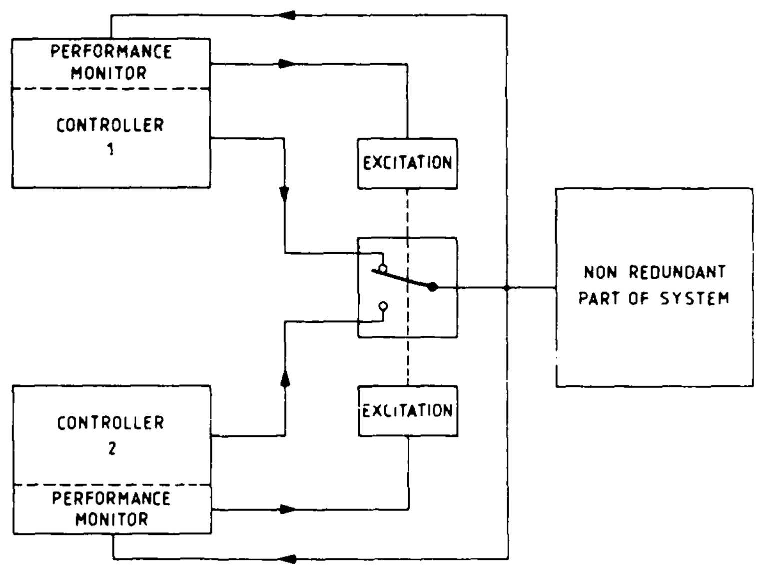


Figure 16 SYMMETRIC SWITCH OVER CONFIGURATION

Figure 17 shows a hiarchically switch-over configuration. This type is particular usefull, when hot-standby subsystems with lower performance levels (gracefull degradation) have been accepted. This switch-over principal can also be used in e.g. the sensor subsystem, described in para 3.5.1 , when it is required that both controllers use identical sensor information.

However, this set-up offers less redundancy in the switch-over capabilities than the solution of the previous paragraph.

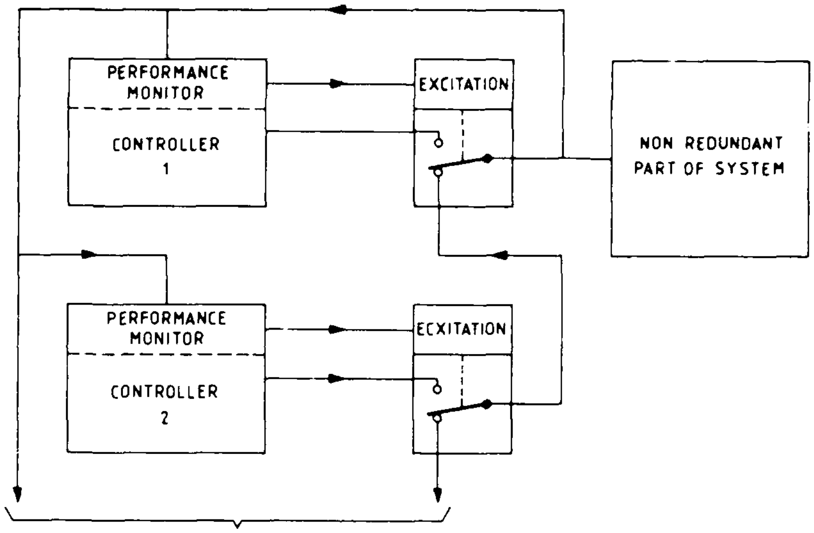


Figure 17 HIERARCHICALLY SWITCH OVER CONFIGURATION

6. FAULT LOCALIZATION

Fault Localization in old fashioned systems has been a relative simple task. With general knowledge of the system, some common sense, a multimeter and some additional human sensing aids for vibration, sound, heat and smell, most failures could be localized relatively easy. Nowadays more and more functions are automated and/or optimized, using increasing amounts of advanced technology. Most of this technology is relatively more reliable but requires, for diagnosis and repair in an operational environment, substantial training of technicians and complex testequipment of increasing complexity and cost. This, combined with a trend to reduce available manpower shows an increasing maintainability problem. The only escape to reduce this problem is to transfer it to the design phase of the equipment, where the designer can add adequate BIT diagnostics into the system. Furthermore Failure Localization routines and procedures and, eventually required external universal test equipment, should be standardized as far as possible.

An identical approach as shown in paragraph 4 , regarding the Performance Monitor, can be used in more detailed level for a good Failure Localization concept. However, an automatically started and sequenced series of off-line diagnostic tests is also acceptable on this level. Also a lower confidence level of the involved detectors may be acceptable, because undetected failures in the Failure Localization level will only cause confusion and elongation of the required repair time, while undetected failures in the Performance Monitor level can cause catastrophic results. Furthermore. BIT being exclusively related to Failure Localization, can make use of comparisons between redundant parts of the system, because the Performance Monitor has indicated already, which of those subsystems is good or bad.

A very strong aid in Failure Localization mechanizations is a memory for the alarmhistory, immediately preceding an (automatic) reconfiguration of a redundant system. To minimize the Mean Time To Repair and required training for technicians, the output of such a failure memory and/or an automatically initiated and sequenced series of tests, should be processed to easy understandable messages or displays to a technician.

An example of a Failure Localization mechanism of a complex servo control loop, as for instance a ship's rudder controller, is given in figure 18 . The basis for the automatic Failure Localization is the detection of a positive or negative servo error, being outside the expectable range, and a failure detection in the feedback channel. It is assumed that the setpoint is rate of change limited, in such a way, that normal operation of the servo loop will never involves large servo errors. An alternative for the setpoint rate of change limitation can be a simple model of the servo loop, indicating when the real servo loop may be operating with a saturated servo error, in which case an alarm is masked.

Provided the feedback function works properly, a certain difference between setpoint and actual parameters will cause a servo error, outside the range being expected in normal operation. This information will be sufficient for a system alarm in the Performance Monitor. For Failure Localization purposes the propagation of this outside range servo error can be traced through the individual blocks of the control system. As long as the output of a block is also over a certain threshold, and corresponds with the sign of the servo error, the failure must be behind this point. A discontinuity in this propagation indicates the location of the failure.

When the controller is part of a redundant configuration, it should be investigated whether the period of time, required for the error propagation through the various blocks, can be completed before the defective controller is disabled. On that moment the status of the various detectors should be frozen for further investigation by a service technician. In a spatially dislocated configuration, where various technologies are involved, part of this information can be used, to indicate to the system operator or watch, what kind of technician has to be sent to which location.

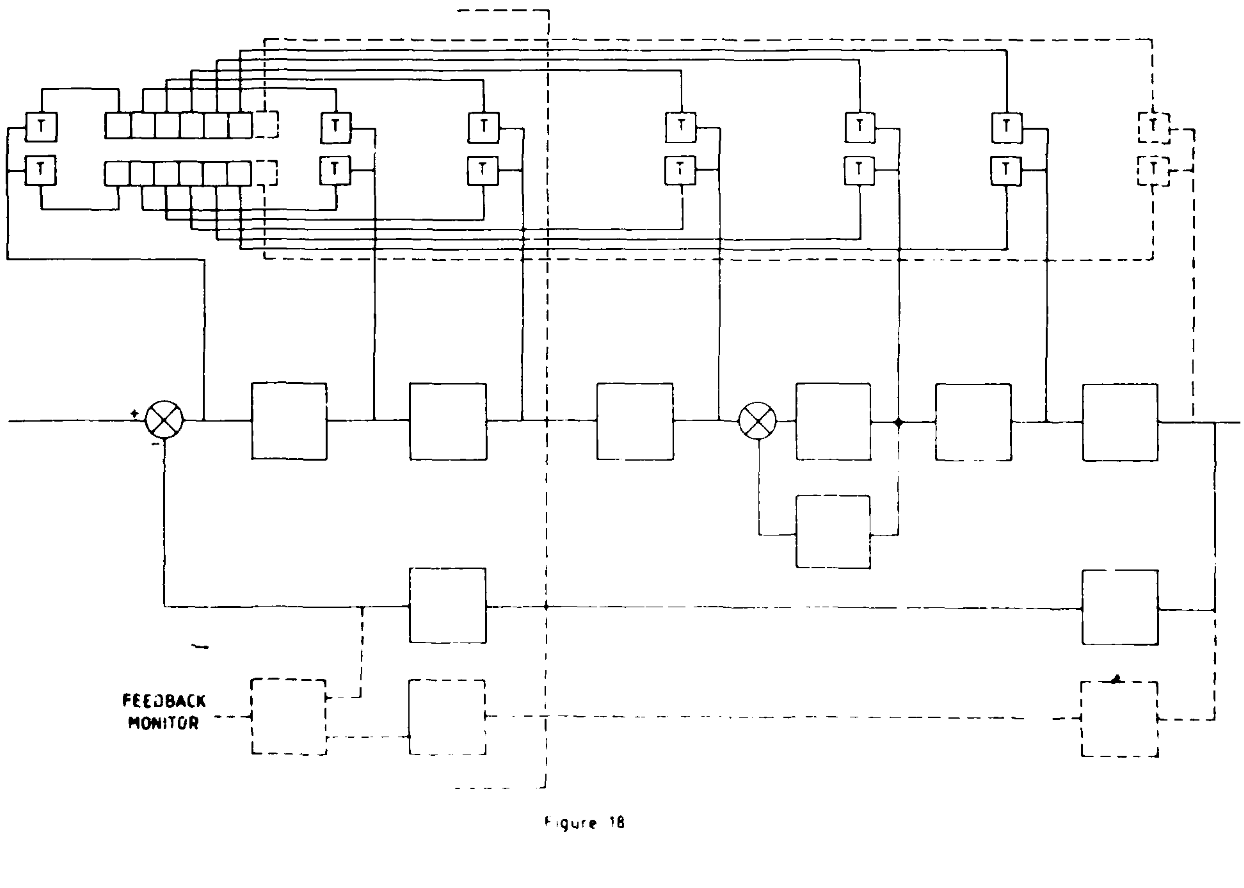


Figure 18

In this concept the feedback function is monitored separately, to avoid potential big differences between setpoint and true actual parameters. These can happen, when a failure in the feedback channel causes a reduced scale factor, a fixed neutral output (open or short circuited connections), or a freeze of a momentary output (tracking process in A/D or S/D converter discontinued). These effects, combined with small setpoint variations , can cause the actual parameter (e.g. the ship's rudder angle) to drift away without adequate timely alarm. The feedback monitor can be based on one or some principles , given under para 3.1 and 3.5 .

7. DEPOT ASSISTED DIAGNOSTICS

In the paragraphs 2 and 7 of this paper it is shown, that Failure Localization with help of BIT is primarily based on an analysis of expectable failures. Furthermore a full 100 % confidence level in a Failure Localization concept may result in unreasonable complexity and/or cost. Automatic location of defective modules or small repair areas is not a goal in itself, but a tool to minimize overall cost of ownership, including all costs related to the availabilty of well trained techniciens on the spot where they are needed and required complex universal testequipment. The other primary goal is to minimize the Mean Time Between Failures.

Because of the reasons mentioned above, additional access possibilities for further diagnostic procedures will be required. So, for instance, the need for conventional testpoints can not be eliminated. Furthermore, especially for investigation of vague or spurious problems by technicians from a depot or manufacturor, a possibility to excecute special diagnostic on-line testroutines will be required. The use of microcomputer development systems and emulators or logic analysers can be very unpractically under operational conditions, especially when rarely occurring strange failure conditions have to be investigated. In those cases it will be very helpfull, when the (PROM programmed) procescomputer has a software programmable breakpoint facility, a (RS 232) interface possibilty for an external terminal or small personal computer, some routines for communication with such a device and some spare Random Access Memory. This set-up enables a special diagnostic testroutine, containing also a (hidden) trigger for the failure to be investigated, can be loaded from the external device into the extra RAM of the procescomputer. In the on-line operational condition, this routine will be started by the software programmable breakpoint and collect a continuously refreshing short history of caracteristic systemdata and information about programflow. This information can be transferred to the external terminal or some memory device on a repetitive basis, or certain events Failure trigger) actually happen. Such a facility is also particularly usefull for acceptance trials, warranty investigations and try-out of small program modifications. The special testroutines for this facility will only be written by people of a depot or the manufacturor. The extra hardware, involved in the procescomputer, can be concentrated on one special printed circuit card, that is only inserted during this type of investigations.

8. CONCLUSION

Provided, that the effects of equipment failures are evaluated in an early phase of the development, adequate Built In Test (BIT) facilities for automatic detection of failures can be incorporated in a system. Based on such facilities a Performance Monitor concept, that will alarm all critical failures in the total system with a high confidence level, can be designed and implemented. Another subset of these facilities can be used to ease Failure Localization, minimizing the Mean Time To Repair (MTTR) and the overall costs, related to the need for availability of well trained technicians and expensive universal testequipment.

The confidence level of a Performance Monitor concept, based on well defined requirements about the conditions where an alarm shall occur, can be calculated. This approach shows critical parts in the design and the effect of additional BIT detectors. In some cases it can be demonstrated, that relative minor changes in the requirements may substantially affect the Performance Monitor concept.

Based on a reliable Performance Monitor, a fail operational redundant controller configuration can reach very high functional Mean Time Between Failures specifications. Using the fact, that on a ship a failure can be repaired restoring the full redundant availability, this can even be true when standard industrial equipment and production techniques are used. However, this assumes that common failure sources, as e.g. caused by environmental conditions, can not occur and adequate Failure Localization BIT takes care of a good repairability.

In such a system the occurrance of a failure leads to an operator alarm, but not to a degradation of functional performance. Such an alarm should lead to a repair procedure, to restore full redundant availability. Only between the occurance of a first failure and the completion of the repair cycle, a second failure may cause a functional system failure. Therefore in those periods increased operator or watch attendance is required and the start of critical operations could be postponed. The need for continuous availability of well trained operators, being able to perform adequate manual control, can be eliminated.

REFERENCES

(1) Richard W. Comerford and Jerry Lyman , "A little extra circuitry on chip and special software lets VLSI test itself chip by chip and board by board" , Electronics / March 10 , 1983

(2) National Transportation Safety Board , Bureau of Accident Investigation Washington, D.C. 20594 , "Aircraft Accident Report - American Airlines, Inc. DC-10-10, N110AA, Chicago-O'Hare International Airport Chicago , Illinois, May 25, 1979

A PRESENT DAY VIEW ON SOFTWARE RELIABILITY

by C. J. Dale
National Centre of Systems Reliability, UKAEA

ABSTRACT

In recent years there has been a trend towards increasing use of software in systems of all kinds, including those whose reliability is a primary concern. This paper examines the ways in which software can affect the reliability of systems, and identifies reasons for the separate treatment of software reliability. It is pointed out that there is a need to engineer reliability into software, exactly as there is for hardware, but that the methods are inevitably different due to the contrasting nature of hardware and software failure processes. Methods for avoiding faults in software are discussed, together with fault removal methods. The need for fault tolerant software for particular applications is also addressed. The state-of-the-art in software reliability assessment and prediction is described, and contrasted with widely accepted methodologies for hardware reliability evaluation. Reasons for the disparities between hardware and software reliability assessment techniques are suggested, and the difficulties of producing overall system reliability assessments addressed. Finally, the paper identifies the capabilities and limitations of present-day methods, identifying topics requiring further research effort, and suggesting the directions that research should take.

INTRODUCTION

Over the past few years there has been increasing use of computers in ship control systems, as indeed there has been in virtually all walks of life. Computers have often been used to perform functions previously carried out by other means. Examples include navigation systems and autopilots. The reasons for using computers in such areas of application include savings in space and weight, speed of operation, enhanced reliability, and cost savings - in addition to which there is often improved performance. Computers have also been used for novel applications, which had previously been regarded as either impracticable or impossible. Examples of such applications include systems which are capable of maintaining a ship's position above a fixed point of the seabed. Thus computers have become, and are certain to remain, critical elements of ship control systems.

Broadly speaking, computers impact on systems reliability in two ways - hardware-induced failures, and software-induced failures. Early computer hardware was notoriously unreliable, so that for many years the software aspect was essentially ignored. However, with the passage of time, computer hardware has become exceptionally reliable, and continues to become even more so. Computer hardware has also become smaller, cheaper and faster - all of which have led to computers being asked to do more complex tasks. This is especially so of real-time systems, which includes the vast majority of ship control applications. More complex tasks require more complex software. These trends towards more reliable hardware and more complex software are both causing software to become the dominant aspect of computer systems unreliability.

This paper concentrates on the software aspect of computer systems reliability, explaining first why it is necessary to consider software reliability separately from hardware reliability, and then discussing ways of achieving reliable software, and of assessing software reliability.

SOFTWARE RELIABILITY

It is sometimes argued that any distinction between software and hardware reliability is unnecessary and artificial, since it is system failures which the user experiences. There are also those who claim that the term "software reliability" is a misnomer - there can be no such thing, since software is either right or wrong. I believe that these arguments arise from a lack of understanding of the software failure process.

A hardware-induced failure is almost certain to be due to some physical change in some component of the system. This may be caused by overstressing of some sort, or be due to an undetected defective component failing within its specified operating environment, or it may be simply a case of a component wearing out. Much hardware theory is based on a belief that any given component will, eventually, fail.

Software, on the other hand, is very much an abstract product, not subject to random physical changes. Software-induced failures are the result of faults in the software itself; these faults (or bugs) lead to failure when particular inputs are encountered. In this sense software reliability is deterministic - the inputs determine whether or not failure will result. It is of course not known in advance which inputs will lead to failure, and the software failure process is usually regarded as a random one, the source of randomness being the stream of successive inputs, each of which may or may not cause the manifestation of a fault.

The above distinction between hardware and software failure processes is fundamental: other differences, such as those listed by Kline and Schneidewind(1), are consequences of it. It is because of this fundamental difference that software reliability needs to be considered separately. Reliability of hardware is achieved by the application of engineering principles and scientific theories. These principles and theories are, in general, completely inapplicable to software, since software faults are essentially caused by human error during production, rather than by some physical change during use. Thus the production of reliable software is crucially dependent upon methods for the avoidance of human error; such methods for software production have come to be known collectively as software engineering.

ACHIEVEMENT OF RELIABILITY SOFTWARE

There are essentially three ways in which software reliability can be achieved: by avoiding the creation of faults in the software, by removing faults which have been created, and by designing the software in such a way that it can tolerate the presence of faults. These three broad strategies can be applied at various stages of the software life-cycle.

Software Life Cycle

In order to facilitate the subsequent discussion, it is necessary to establish a framework for the software life-cycle, as it relates to the system life-cycle. This is shown in Figure 1.

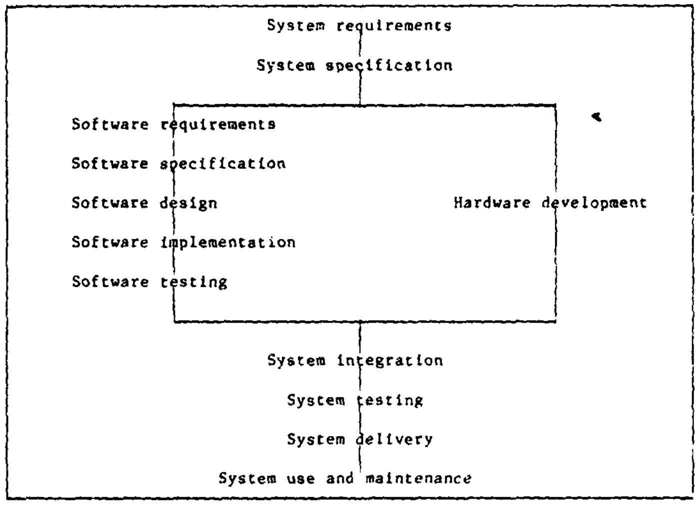


Figure 1. Software and System Life-Cycles

The life-cycles shown in Figure 1 should not be taken as definitive. In particular, the separation between hardware and software development may not be as great as indicated, and the strict sequentiality implied by Figure 1 is not necessary.

Software faults can arise both within and between the various stages of the life-cycle. Faults within stages include such things as inaccurate requirements, ambiguous statements in a specification, and human errors in design or implementation, where the human concerned intended to do the correct thing but has inadvertently done otherwise. Faults between stages are often misunderstandings, such as misinterpretation of the specification by a designer. In such a case the designer may correctly create a design for the wrong software. Fault avoidance techniques thus need to address faults within stages and faults between stages.

The lesson of history is that no matter how hard we try, we shall still make mistakes Thus there is a need to apply fault removal techniques to isolate and remove those faults which creep in, despite the efforts to prevent them. Fault removal can be applied at all stages of the life-cycle, including such activities as checking a specification for ambiguities. It is worth emphasising that great savings can be made by finding faults as soon as possible after their creation, as is demonstrated by Boehm[2]. It is also the case that fault removal is no substitute for fault avoidance: as with hardware, software has to be engineered to be reliable; reliability cannot be tested in.

For many applications, a combination of fault avoidance and fault removal techniques produces software of sufficient reliability. If this is not the case, however, recourse can be made to methods of fault tolerance, which means that the software itself is capable of detecting an erroneous internal state and doing something about it before system failure results.

Fault Avoidance Methods

Potier et al[3] report a study in which more than 60% of all faults originated prior to the implementation (or coding) stage. It is thus essential that fault avoidance techniques are applied at all stages of the development, and not just during the programming itself. Since software faults arise from human errors, the control of factors which affect the occurrence of human errors will reduce the fault content of software.

According to Rzevski[4], there are five classes of factors which affect the occurrence of human errors: these are discussed below. The choice of development methods which assist control of these factors can be expected to produce software which is relatively fault free.

System Complexity. The inability of humans to deal with over-complex entities without committing numerous mistakes is well documented, for instance by Miller[5]. There is a number of software design methodologies which are all aimed at containing the complexity of software. Many of these organise the software in a hierarchy of reasonably independent software modules (e.g., Yourden and Constantine[6], Myers[7]). An alternative approach is due to Jackson[8]. There is thus a range of methods available for control of system complexity, which are most readily applied at the design stage.

Task Complexity. Experience shows that the frequency of human errors increases if tasks are too simple (and thus boring), or too complex. This topic has had very little coverage in the literature, and tends to receive insufficient attention from software development managers. A useful rule of thumb which can be applied at the implementation stage is that no individual should be expected to deal with more than one module at a time. In general, tasks should be clearly defined in terms of inputs, outputs and activities to be performed.

Resources. The quality and characteristics of the resources made available for system development can have a significant effect on the occurrence of human errors. Resources include languages, guidelines, manuals, standards, computer aids and tools applicable to all phases of software development. There is a very large number of tools and other resources available, and it is an important managerial task to select the best and most appropriate tools for the job in hand.

The language which is chosen for a particular application may affect considerably the avoidance of faults. It should be borne in mind that it is now possible to express specifications and designs, as well as programs, in formal languages.

Guidelines and standards are very powerful aids to the organisation and management of software production. Company standards are often preferable to national or international ones, because they can be tailored to the particular organisation and can more readily be modified and updated.

There has in recent years been an explosion of activity in the field of computer aids and tools for software production. The problems here for managers are posed not only by the sheer number of tools and the difficulty of deciding which are the appropriate ones, but also by the extreme difficulty of assembling an integrated tool set to deal with the whole of the software development. It is to be hoped that the development in the near future of Integrated Project Support Environments (IPSEs) will alleviate these problems. In the meantime, the Department of Trade and Industry has published a very useful guide to available tools for application to large real time systems[9], covering tools for project control, requirement specification, the design process, verification, validation and testing, and version and configuration control.

Human Factors. Inevitably, the occurrence of human errors is influenced by the characteristics of the personnel involved in software development. While there is a large body of knowledge on the relationship between human factors and the frequency of errors by operators of machinery, the effects of human factors on design errors such as those which lead to software faults have not been studied extensively. There is evidence[4], however, that the occurrence of human errors is minimised if software production personnel are equipped with knowledge and skills related to systematic design methods, and if their attitude is egoless, thorough, and self disciplined. It is also an important skill to be able to identify simple solutions to complex problems. Thus the control of human factors is by selection and training of personnel.

Environmental Factors. Reliability of software can be affected by the physical, social and psychological aspects of the environment in which systems are created. Examples of possible influencing factors are noise level, team morale, and criteria for promotion. An important reference in this area is the book by Weinberg[10].

Fault Removal Methods

As has already been pointed out, fault avoidance techniques used in isolation will not produce sufficiently fault-free software. It is thus essential to carry out fault removal activities in addition. Historically, fault removal consisted simply of testing the software in an ad hoc manner. More recently, it has been recognised that the longer a fault remains, the more expensive it is to eliminate[2], and fault removal activities are now applied at all stages of the software life-cycle.

Inspections. Substantial improvements in programming quality and productivity can be obtained through the use of formal inspections of design and of code. The chief objective of the inspection process is to find faults. For the sake of clarity, the inspection process will be described as it would be applied between design and implementation. As Fagan[11] points out, inspections can be applied throughout the development cycle of software.

The design inspection team should be led by a moderator, and include the person who produced the design, the person who will implement that design, and the person who will eventually test the program produced. Preliminary to the inspection itself there is an overview meeting at which the designer describes the overall area being addressed and the specific area he has designed in detail. This is followed by individual preparation so that each member of the team has an understanding of the design prior to the inspection meeting. The inspection meeting itself consists of the implementor describing how he will code the design. Other members of the team raise questions during the implementor's discourse, which are pursued to the point of a fault being identified. After the inspection meeting all faults noted are resolved by the designer. It is the moderator's responsibility to ensure that all problems have been resolved.

The details of the above description will inevitably vary for application of inspections to other stages of the software life cycle, but the underlying principle of organising a formal detailed discussion between interested parties with the objective of identifying faults is applicable throughout the software life-cycle.

Formal Methods of Software Development. These methods are much more than fault removal methods, addressing as they do the whole process of software development. The reason for their inclusion here is their great power at removing faults, especially in the early stages of the life-cycle.

The aim of formal methods is to make available concise, unambiguous specifications and designs, which can then be derived into programs which are correct with respect to their specifications. This is based on a view that only a formal mathematical basis can bring order into software development.

Probably the most highly developed formal method is the Vienna Development Method (VDM)[12], originally developed by IBM. VDM employs a formal mathematical notation, and the basic procedure is to write a rigorous specification of the software which is then developed into a derived correct program. The final stage is to examine the performance of this program and improve its efficiency where necessary.

Users of VDM report[9] that the act of constructing a VDM specification reveals inconsistencies, ambiguities and incompleteness in initial requirements, and that the specification is highly effective as a medium for resolving such defects with the customer.

Testing and Debugging. Testing is the oldest software fault removal method, and over the years a number of highly sophisticated testing techniques have been developed: many of these are described in the book by Myers[13]. The problem with testing is that it can only take place after the software has been implemented - this is why fault removal techniques such as inspections are necessary in addition to testing.

Ideally, one would like to test a piece of software to find all of its faults. This is in general impossible, as will be shown below. In response to this situation, two broad families of testing techniques have grown up, known as black-box and glass-box testing.

Black-box testing involves deriving sets of input test data from the specifications, and ascertaining whether the correct outputs are obtained. The testing is thus independent of the way in which the software has been written. To exhaustively test software in this way would involve submitting all possible sets of input values to the program. If a given program has six inputs, each of which is a 32-bit number, the number of possible sets of input values is 2^{192}, which is greater than 6×10^{57} - and this for a program which may be very small. Exhaustive testing is thus impossible, and the aim must be to choose representative input sets in such a way that the return on investment in testing is maximised, in terms of faults removed.

Glass-box testing is based instead on a knowledge of the software itself. Exhaustive glass-box testing would involve choosing sets of input values to ensure that every possible path through the software is executed. Again, it is not difficult to devise a ten-statement program with 10^{14} unique paths through it, making exhaustive path testing a practical impossibility in most cases. The various glass-box techniques all aim for some sensible compromise somewhat short of path testing. The simplest (and least effective) of these techniques is statement coverage, which means ensuring that every statement is executed at least once during testing.

Testing is commonly perceived to have two main purposes. One is to find as many faults as possible so that they can be removed, thus improving the reliability. The second is to demonstrate that the software is sufficiently reliable. These two aims are, unfortunately, in total conflict. For this reason, it is probably best to separate the two activities, dealing first with the fault removal aspect, and secondly with the demonstration aspect.

Fault tolerance

The utilisation of protective redundant elements of a system is known as fault tolerance. In the case of hardware, fault tolerance may simply be a case of redundancy, where two identical components are employed where one is sufficient, in order to protect the system against failure of one. The decision on whether to employ redundancy is a case of trading off the cost of additional components against the benefits of enhanced reliability.

As far as software is concerned, the utilisation of a second identical piece of software is of course futile, because any faults present will be in both copies, and will lead to simultaneous failure of both when particular input is encountered. Thus fault tolerance in software can only be achieved by employing dissimilar copies of software. There are two principal methods of achieving fault tolerance in software, each of which is discussed below.

N-Version Programming. This strategy for software fault tolerance was suggested by Avizienis and Chen[14], and consists essentially of a number of independently programmed implementations of the same specification. The separate software versions are run separately, and their outputs compared. In the event of disagreement, some form of voting system or other pre-determined strategy is used to decide which is the preferred result. As well as the cost involved in replicating software development, there are two major technical problems. The first is that faults introduced at or before the specification stage are likely to be present in all versions of the software, and thus will not be detected by the voting mechanism. The second difficulty is that it has not been established to what extent the failures of independently-written versions of software will be independent. These problems make it impossible, at the present time, to quantify the improvements in reliability likely to be achieved from the use of n-version programming. It can therefore only be recommended at the moment for use in applications where the benefits of improved reliability are considerable, such as in systems whose failure may lead to loss of life.

Recovery Block. This second software fault tolerance strategy is due to Randall[15], and differs from n-version programming in that the fault tolerance is incorporated in a single program. A recovery block consists of a section of code in some programming language, together with an acceptance test and a number of alternative algorithms. The purpose of the acceptance test is to detect processing errors during the execution of the section of code. If the result is acceptable, then processing continues with the next block. If it is unacceptable, all variable values are restored to their values before the section of code was executed, an alternative algorithm is executed, and the acceptance test repeated. This process can be repeated until all the alternates are used up, in which case the recovery block has failed. The problems identified earlier with respect to n-version programming apply equally to recovery blocks, and so the same comments apply.

ASSESSMENT OF SOFTWARE RELIABILITY

State-of-the-Art

Over the past fifteen years there has been a great deal of work aimed at estimating and predicting software reliability. The bulk of this work is based on a scenario of a piece of software undergoing testing, and having changes made to it whenever a failure is observed, with the intention of fixing the fault which led to the failure. The various software reliability prediction models make use of a knowledge of the times at which failures have occurred to estimate the current failure rate, and predict the failure rate at future times.

The best known of these models are due to Jelinski and Moranda[16], Littlewood[17], Littlewood and Verrall[18] and Musa[19]; software reliability evaluation methods generally have been surveyed by Dale[20]. The problem with all of these models, and the reason for the existence of a number of competing models, is that none of them is capable of predicting to an accuracy which is considered adequate. Recent work by Keiller et al[21] indicates that, as far as failure rate estimation and short-term prediction are concerned, some prediction methods perform quite well some of the time, but none of them performs well all the time. Their conclusion is that there is no universally best model, and they recommend trying several models and examining the quality of prediction obtained from each. Tools are available to do this for a limited range of models.

Software v Hardware Reliability Assessment

Over the years that hardware reliability has been seriously studied, a number of analytical reliability assessment techniques have become widely accepted and used, and even standardised. Some examples include Fault Tree Analysis (FTA), Failure Mode Effect and Criticality Analysis (FMECA), and the American Military Standard 217 approach to reliability prediction for electronic systems (MIL217); all of these are described by O'Connor[22]. FTA is a technique for analysing how particular system failure events can happen, and FMECA can be used to analyse the consequences for a system of the failure of various components. Both of these techniques are qualitative in nature, but both have been successfully extended to give quantitative estimates of the various failure rates and other reliability measures. MIL217 is by its very nature a quantitative technique, which makes use of a large data base of component failure characteristics to predict the failure rates of electronic systems, based on a knowledge of the design.

These techniques have in common that they can be used long before any hardware is built. Thus they are useful for influencing design decisions and development strategy in order to ensure that reliability requirements can be met. The software reliability prediction techniques discussed above can be used only after the software has been written, and even then they are not sufficiently well developed to permit prediction of reliability during customer use from a knowledge of reliability during testing.

A consequence of this disparity between hardware and software reliability prediction techniques is that the current ability to predict systems reliability at an early stage of the life-cycle is entirely dependent upon the ability to predict hardware reliability. As electronic hardware becomes more reliable, and as the software content of systems increases, so the ability to predict systems reliability will decrease - unless methods are developed enabling the assessment of software reliability at an earlier stage of the life-cycle than is currently possible.

The reasons for the disparity are many, and include the fact that hardware reliability is a much older and more mature discipline. Another important reason is that hardware is built from commonly-used components whose failure chracteristics are often well-understood, whereas the overwhelming tendency with software development is to create the whole system from scratch, with little or no use of already existing software. In addition, I would argue that failure of hardware is inherently more predictable. Component failures and their consequences for the system are, in general, readily anticipated, and account for the majority of hardware-induced system failures. If, on the other hand, particular software failure modes were anticipated, they could be tested for and any faults found eliminated. Thus every software-induced failure is an unexpected event - hardware-induced failures, by contrast, are in general events which were expected to happen sooner or later.

CONCLUSIONS AND RECOMMENDATIONS

In the past few years a host of software development methodologies, methods, techniques and tools have become available, many of which are of great potential benefit to the development of reliable software. Unfortunately there is virtually no information available to enable an intelligent decision to be made as to which software development strategy is likely to produce software with a given target reliability. There is also currently great difficulty in assembling a cohesive set of techniques to provide for the whole of software development; this problem is likely to be reduced by developments expected to take place over the next few years.

As far as reliability assessment of software is concerned, it is now possible to estimate the failure rate of a piece of software with reasonable accuracy. This can, however, only be done once the software has been written, and problems of modifying estimates to take into account the likely variation in failure behaviour in different usage environments have yet to be addressed.

There are a number of topics in need of research to address the problems identified above. Of particular importance is the need to research, and where possible quantify, the effectiveness of the various software development techniques. The information provided by this research could be used to decide which techniques and tools should be employed to achieve particular reliability and other goals. This should in turn produce better estimates of timescales and costs of software development, as well as avoiding wasteful reworking when it is discovered - too late - that reliability is inadequate. In addition, management will be more willing to invest in tools if they are given a quantified idea of the likely returns.

This research will also greatly assist the development of quantified reliability assessment methods for use throughout the software life-cycle. Such methods need to be developed to provide management with information as to whether and when targets are likely to be met. There is currently no quantitative software reliability assessment method which can be used early in the life-cycle. The best that is available is the qualitative approach reported by Daniels[23], which is based on a checklist of questions to be asked by a software reliability assessor. This approach provides a starting point for work in the area.

REFERENCES

(1) M. B. Kline and N. F. Schneidewind, "Life Cycle Comparisons of Hardware and Software Maintainability", Proceedings of the 3rd National Reliability Conference, 1981.

(2) B. W. Boehm, "Software Engineering Economics", Prentice Hall, 1981.

(3) D. Potier, J. L. Albin, R. Ferreol and A. Bilodeau, "Experimenting with Computer Software Complexity and Reliability", Proceedings of the 12th Annual International Symposium on Fault-Tolerant Computing, 1982, p381-389.

(4) G. Rzevski, "Recent Advances in Software Reliability Methods", Proceedings of the 3rd National Reliability Conference, 1981.

(5) G. A. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information", Psychological Review, March 1956, p81-97.

(6) E. Yourden and L. L. Constantine, "Structural Design: Fundamentals of a Discipline of Computer Program and Systems Design", Prentice Hall, 1979.

(7) G. J. Myers, "Composite/Structured Design", Van Nostrand, 1978.

(8) M. A. Jackson, "System Development", Prentice Hall, 1983.

(9) "The STARTS Guide: Software Tools for Application to large Real Time Systems", Department of Trade and Industry, 1984.

(10) G. M. Weinberg, "The Psychology of Computer Programming", Van Nostrand, 1971.

(11) M. E. Fagan, "Design and Code Inspections to Reduce Errors in Program Development", IBM Systems Journal, 1976, p182-211.

(12) D. Bjorner and C. B. Jones, "Formal Specification and Software Development", Prentice Hall, 1982.

(13) G. J. Myers, "The Art of Software Testing", Wiley, 1979.

(14) A. Avizienis and L. Chen, "On the Implementation of N-Version Programming for Software Fault-Tolerance During Program Execution", Proceedings of COMPSAC, 1977, p149-155.

(15) B. Randell, "System Structure for Software Fault Tolerance", IEEE Transactions on Software Engineering, 1975, p220-232.

(16) Z. Jelinski and P. B. Moranda, "Software Reliability Research", in Statistical Computer Performance Evaluation, W. Freiberger (Ed), Academic Press, 1972, p465-484.

(17) B. Littlewood, "A Bayesian Differential Debugging Model for Software Reliability", Proceedings of Workshop on Quantitative Software Models, 1979, p170-181.

(18) B. Littlewood and J. L. Verrall, "A Bayesian Reliability Growth Model for Computer Software", Applied Statistics, 1973, p332-346.

(19) J. D. Musa, "A Theory of Software Reliability and its Application", IEEE Transactions on Software Engineering, 1975, p312-327.

(20) C. J. Dale, "Software Reliability Evaluation Methods", British Aerospace Dynamics Group Report ST26750, 1982.

(21) P. A. Keiller, B. Littlewood, D. R. Miller and A. Sofer, "Comparison of Software Reliability Predictions", Proceedings of the 13th Annual International Symposium on Fault-Tolerant Computing, 1983.

(22) P. D. T. O'Connor, "Reliability Engineering", Heyden, 1981.

(23) B. K. Daniels, "Software Reliability", Reliability Engineering, 1983, p199-234.

THE CONTROL SYSTEM TEST VEHICLE

by Delbert C. Summey

Vehicle Technology Division, Code 420
Naval Coastal Systems Center

ABSTRACT

The Control System Test Vehicle (CSTV) is a 1/12-scale 688 autonomous submarine for hydrodynamics and controls research. Designed for System Identification application, the CSTV performs preprogrammed maneuvers while gathering a large amount of dynamic response data to gain insight into the equations of motion of submarines as well as vehicle control algorithms. The physical characteristics of the CSTV are discussed along with the on-board instrumentation system. The vehicle operation is described and dynamic test data is presented for selected dynamic maneuvers.

INTRODUCTION

Techniques for the design of high performance submarines have been investigated under the direction of the Advanced Submarine Control Program (ASCOP), a Naval Sea Systems Command (NAVSEA) Advanced Development R&D Program. NCSC played a key role in the ASCOP program through the design, development, and operation of the Control System Test Vehicle (CSTV).(1) The CSTV is an autonomous vehicle with provisions for different external geometry configurations for submarine hydrodynamic and control systems research. The vehicle is capable of performing a full spectrum of submarine maneuvers, including emergency recovery maneuvers, and is instrumented to permit System Identification analysis of control system input and the resulting vehicle motion output data.

System identification is a methodology for the identification of estimation of parameters and structures of dynamic systems based on observations of system inputs and outputs.(2,3) To use System Identification, either a model or the actual full-scale craft is operated under conditions that excite motions similar to those for which mathematical models are to be obtained. The maneuvers must be very violent in many cases in order to make the hydrodynamics observable in the data. Large quantities of highly accurate trajectory data are required to successfully apply the System Identification technology to the development of improved submarine equations of motion. The cost, limited availability, safety limitations on extreme maneuvers, and the impracticality of modifying the hull geometry and control surface configurations precluded the use of a full-scale submarine to obtain the System Identification data.

The initial specifications for the CSTV and support equipment were generated by NCSC in cooperation with the David Taylor Naval Ship Research and Development Center, Naval Ship Engineering Center, Naval Underwater Systems Center, and NAVSEA. Final design and construction of the vehicle was accomplished by Lockheed Missile and Space Corporation. The navigation, guidance, control and communication subsystems as well as all vehicle computer software were developed by NCSC and integrated into the vehicle.

VEHICLE PHYSICAL CHARACTERISTICS

The design of the CSTV provides for the basic geometry variations illustrated in Figure 1. The modular sail may be positioned in any of five different locations or removed completely. The modular bow planes may be positioned on the bow or utilized as sail planes or again, removed. A parallel mid-body provides the capability to change the overall length to diameter ratio of the vehicle. The vehicle is also equipped two different tail configurations and is fitted with a scaled version of the standard propellor while offering the potential to test other propellor configurations. The sail and each control surface are instrumented with strain gauges to provide valuable force and moment data for hydrodynamic analysis.

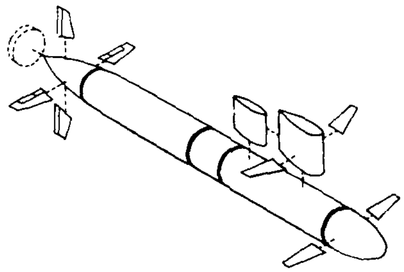


Figure 1. CSTV Geometry Variability.

Physically, the CSTV is a 1/12-scale model of the SSN 688 Class submarine (Figure 2), and consists of a pressure hull, nose and tail structures, and a sail assembly. The all-aluminum pressure hull, 30 feet in length and 33 inches in diameter, is constructed in four sections and contains an inertial measurement system, computer, tape recorder, batteries, trim tanks and related equipment. Propelled by a 25-horsepower electric motor powered via silverzinc batteries, the vehicle's endurance is seven hours below 10 knots or one hour at maximum speed. The maximum operating depth of the 8500-pound vehicle is 300 feet with survivability to 1200 feet. The CSTV's displacement, centers of buoyancy and gravity, mass moments of inertia, and the moments associated with flooding and blowing the ballast tanks are within specified limits of the scaled SSN 688 properties. A detailed description of some of the internal components is provided in Figure 3.

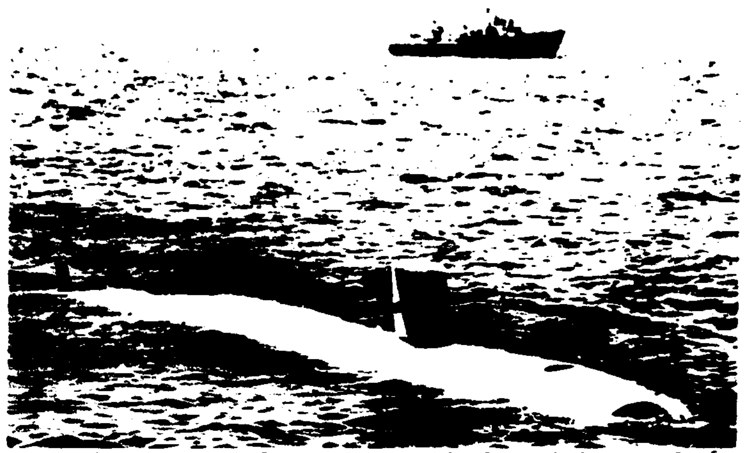


Figure 2. The Control System Test Vehicle With Support Craft.

VEHICLE INSTRUMENTATION

The mission of the CSTV is to perform controlled maneuvers while sensing and collecting data. The CSTV instrumentation package is central to this function and consists of a Control and Recording System, Inertial Measurement System, Control and Display System, and an Acoustic Positioning and Telemetry System. Figure 4 depicts the interrelationships of the CSTV instrumentation package.

The central processor of the Control and Recording System is the FORTRAN programmable AN/UYK-19 ROLM Model 1664 Computer that runs under a real-time, multi-task operating system. During operation, the computer receives 64 channels of analog data as well as parallel digital data from the Inertial Measurement System and the Acoustic Positioning and Telemetry System. The computer sends digital commands to the magnetic tape unit, control surface actuators, and 40 other onboard functions. Approximately 100 variables are recorded onboard the CSTV at a 10 Hz data rate.

The Inertial Measurement System, a slightly modified version of the Honeywell Aerospace Division Advanced Tactical Inertial Guidance System, measures the dynamic motions of the CSTV and may be aligned at dockside or at sea. The position, attitude, and velocity of this two nautical mile-per-hour system are blended with the paddle wheel longitudinal water relative velocity, the two depth gauges at the nose, and the range measurement to the support craft using an 18-state Kalman filter algorithm to produce accurate navigation and other vehicle state information. Estimates are made for latitude, longitude and depth error corrections; north, east, and down inertial velocity corrections; three platform tilt angle corrections; three accelerometer biases; three gyro biases; and north, east, and down water currents. Using these corrections, improved position, velocity, and attitude information is achieved.

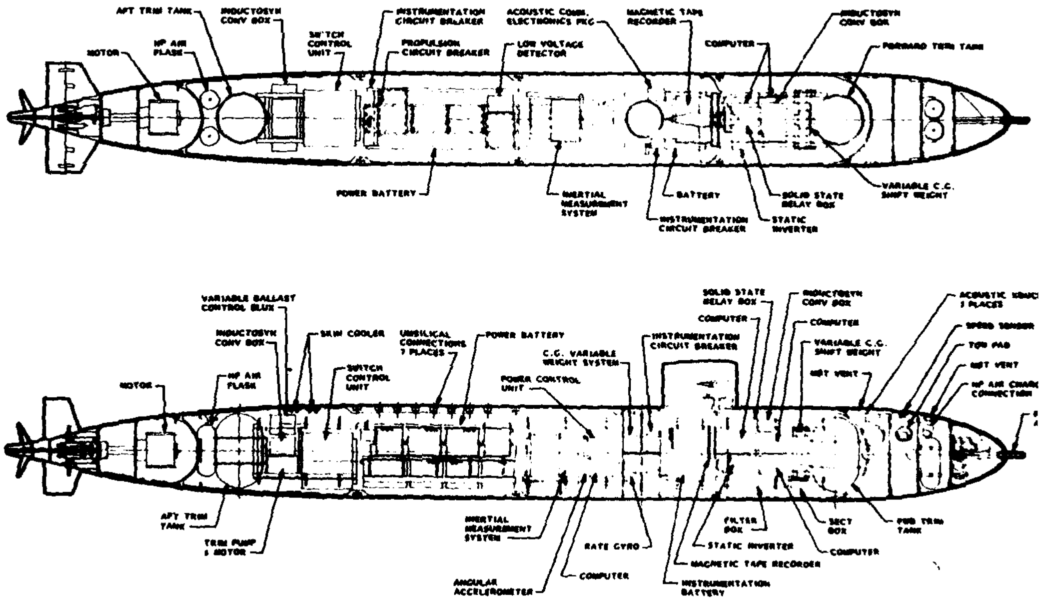


Figure 3. Internal Components Of The CSTV.

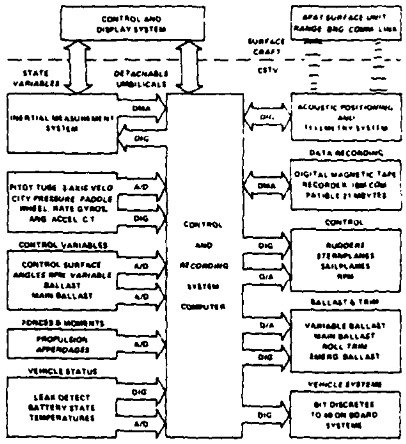


Figure 4. CSTV System Schematic.

Transformation between vehicle body axes and the local level, together with the square root nature of the range measurement, required the use of an extended Kalman filter. This was achieved using a second order Runge-Kutta integration method with a time step of 0.8 seconds providing an intermediate estimate of the state corrections at each half step or every 0.4 seconds. The overall navigation, guidance and control system logic is depicted in Figure 5.

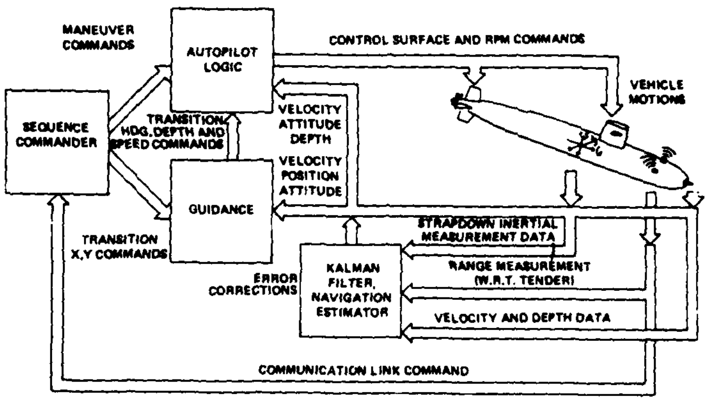


Figure 5. Navigation, Guidance, And Control System Logic.

During initialization and GO/NO GO system checks, the CSTV is connected through umbilicals to the Control and Display System. The Control and Display System provides a means for the operators to perform prelaunch operations, including bring up CSTV power distribution systems with shipboard power; initialization, alignment and testing of the Inertial Measurement System, initialization and auto-test of the Control and Recording System computer, sensors and activator subsystems; direct activation of some ballast and trim functions; and transfer to CSTV onboard power.

When the Control and Display System umbilicals are detached from the CSTV, the Acoustic Positioning and Telemetry System provides the operators with a low data rate command and status link. A video display, generated from range data, presents the CSTV range in the reference frame of the support craft. Eight-bit commands may be sent to the vehicle via an HP-85 microcomputer keyboard aboard the support craft, some of which are interrogations requiring a reply from the CSTV. On board the CSTV, range to the support craft is determined, and the command is decoded and presented to the Control and Recording System computer. The Control and Recording System computer sends the desired eight-bit reply back to the support craft using the onboard Acoustic Positioning and Telemetry System transmission link. Transmission intervals are 6.56 seconds at both the support craft and CSTV, and transmissions at the CSTV follow topside transmissions by 3.28 seconds. The maximum range of the system is 4800 yards.

VEHICLE OPERATION

An aluminum sled is used to tow the CSTV from dockside to the operating area (Figure 6). The tow boat, which acts as the operational control center, is anchored upon reaching the operating area, and the CSTV GO/NO GO system checks are performed while the Inertial Measurement System is being aligned. Once launched and ballasted, the CSTV is towed to a point approximately 1000 feet from the support craft and started by an acoustic command via the Acoustic Positioning and Telemetry System link from the support craft. A preprogrammed sequence of guidance controller commands is given to the auto-pilot to achieve successive desired combinations of position, heading, attitude and speed, or to perform a prestored maneuver sequence within the four maneuver boxes around the support craft (Figure 7). The vehicle navigation system tracks the vehicle to see if dynamic maneuvers cause the vehicle to leave the boundaries of the maneuver boxes; if the boundaries are broken, the maneuver is automatically terminated and the vehicle returns to the center of the maneuver box to await a command to begin another maneuver sequence. Automatic transition from one maneuver box to another is insured by the vehicle control system. During submerged operations, conformance with each preprogrammed maneuver sequence is monitored via the Acoustic Positioning and Telemetry System.

To date the CSTV has been used to gather data on more than 40 miles of autonomous operations during systems development tests and System Identification maneuvers providing a wealth of dynamic data for investigation and analysis. Examples of this data are provided in Figures 8, 9, and 10.

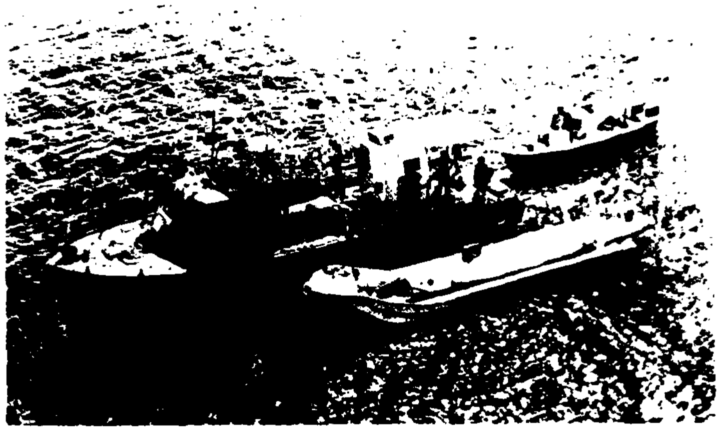


Figure 6. CSTV Sled, Operational Craft, And Support Boat.

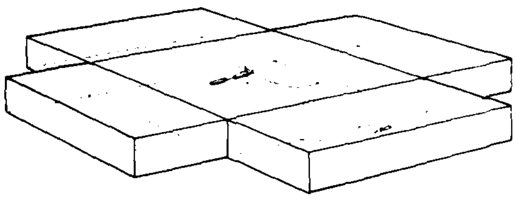
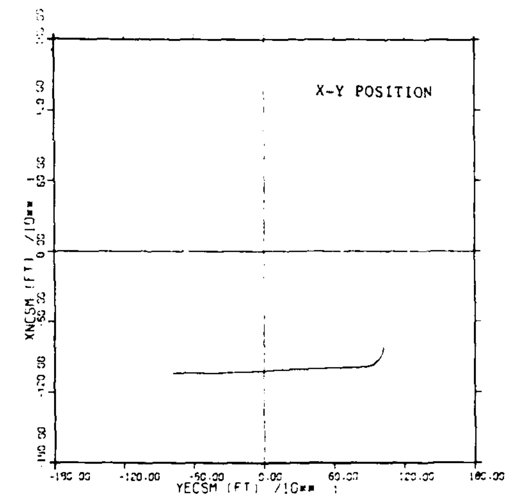


Figure 7. CSTV Operational Area.

The data presented in Figure 8 corresponds to a vertical maneuver in which the vehicle stern planes are commanded to an angle until a given pitch angle is achieved, then the stern plane is reversed until half the same negative pitch angle is achieved. Once this negative pitch angle is encountered the autopilot returns the vehicle to original track and depth and the maneuver is repeated.

Figure 9 provides dynamic response data for a modified lateral maneuver. The rudder angle is deflected to a prescribed angular position until a given heading is achieved; then, the rudder angle is reversed until the negative heading angle is achieved. Once the final heading angle is achieved the autopilot is activated to return the vehicle to depth and track and the maneuver is repeated. The maneuvers is actually a modified lateral maneuver it that the longitudinal autopilot was seeking to maintain a fixed pitch angle during the maneuver to excite cross-coupling effects.



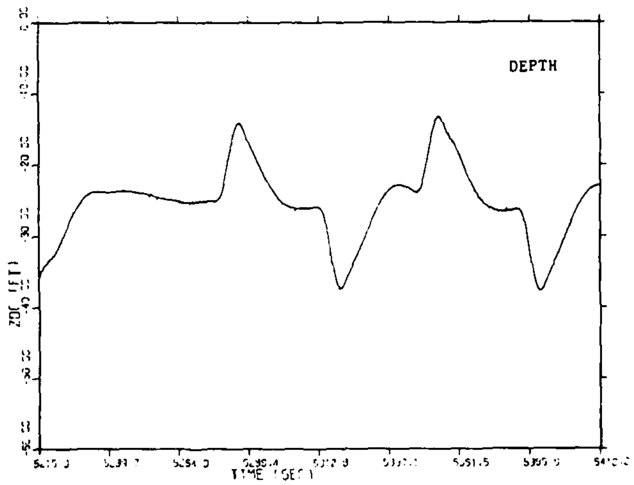


Figure 8. X-Y Position, Depth, And Pitch, Roll, And Heading Angle Histories For A Typical CSTV Vertical Maneuver (1 of 3).

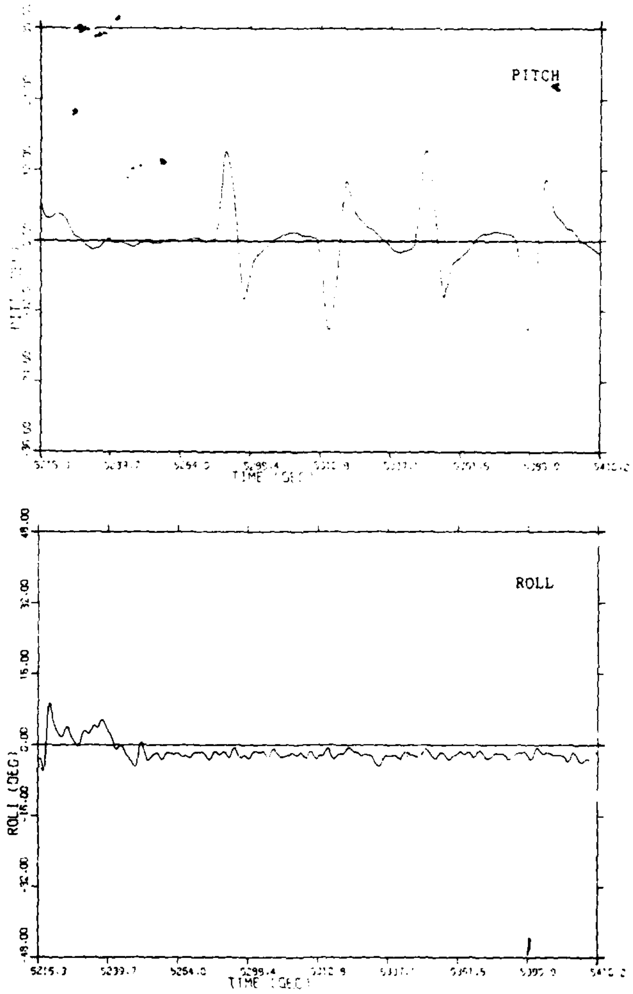


Figure 8. X-Y Position, Depth, And Pitch, Roll, And Heading Angle Histories For A Typical CSTV Vertical Maneuver (2 of 3).

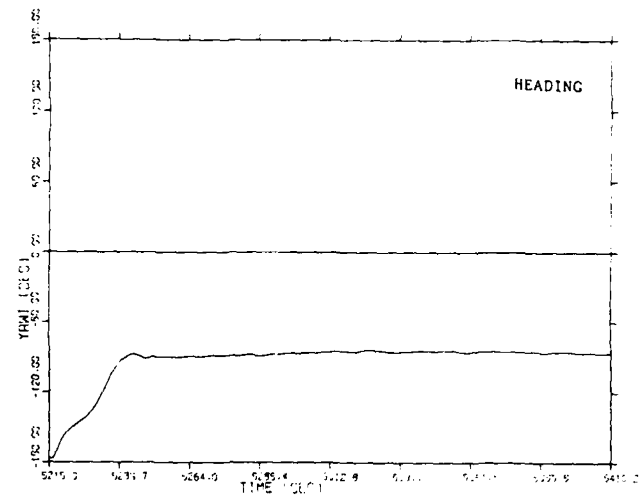


Figure 8. X-Y Position, Depth, And Pitch, Roll, And Heading Angle Histories For A Typical CSTV Vertical Maneuver (3 of 3).

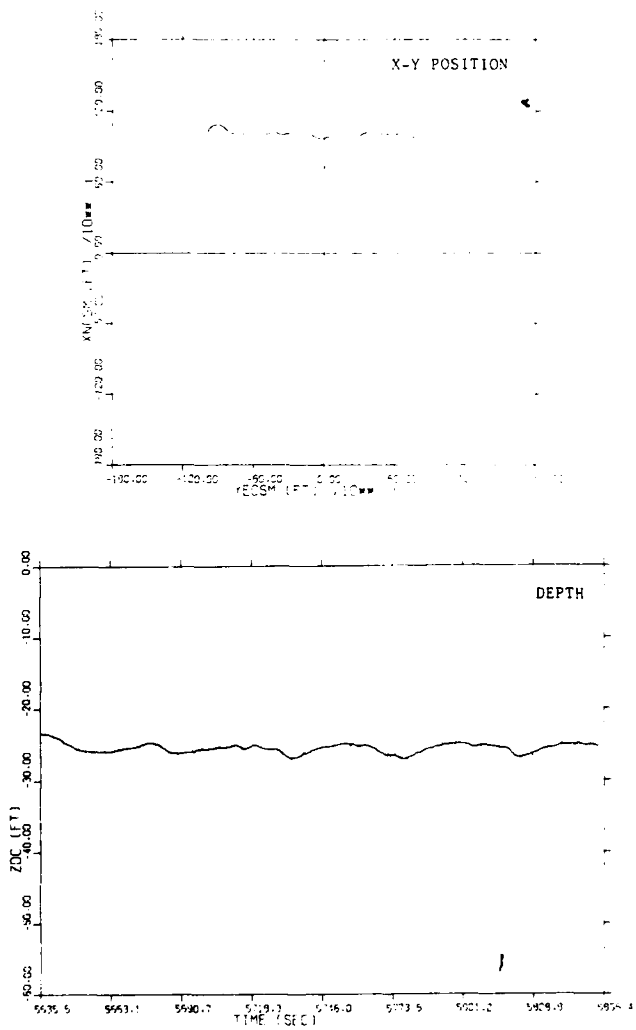


Figure 9. **X-Y Position, Depth, And Pitch, Roll And Heading Angle Histories For A Typical CSTV Horizontal Maneuver (1 of 3).**

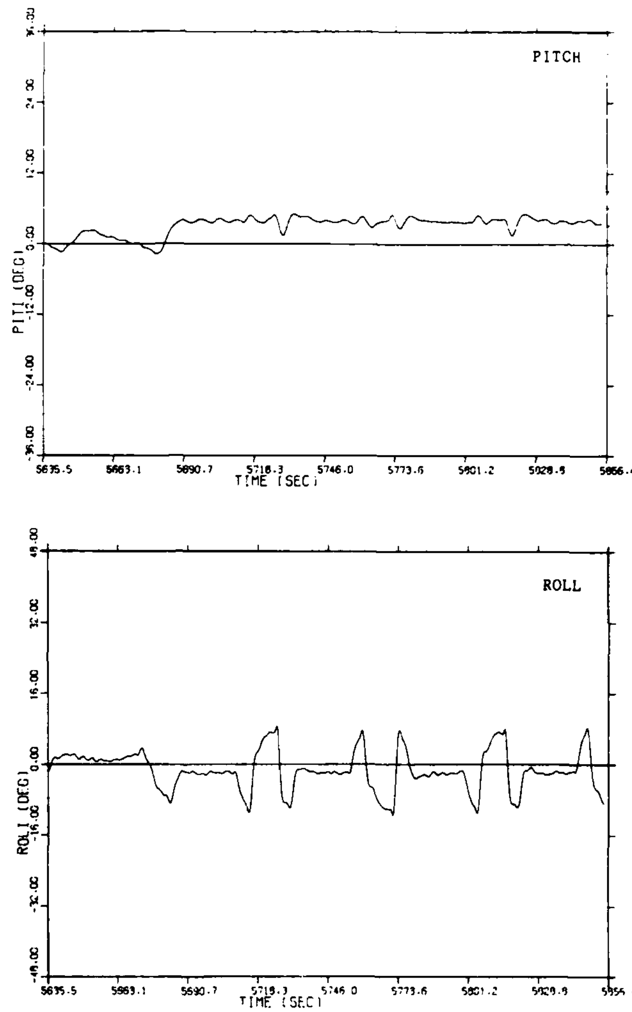


Figure 9. X-Y Position, Depth, And Pitch, Roll And Heading Angle Histories For A Typical CSTV Horizontal Maneuver (2 of 3).

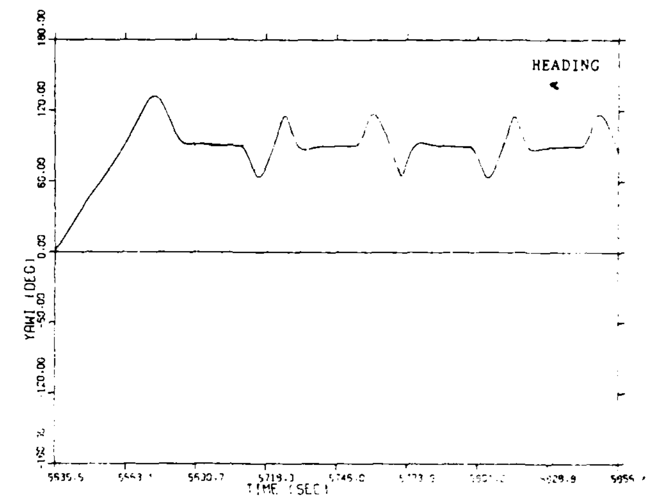
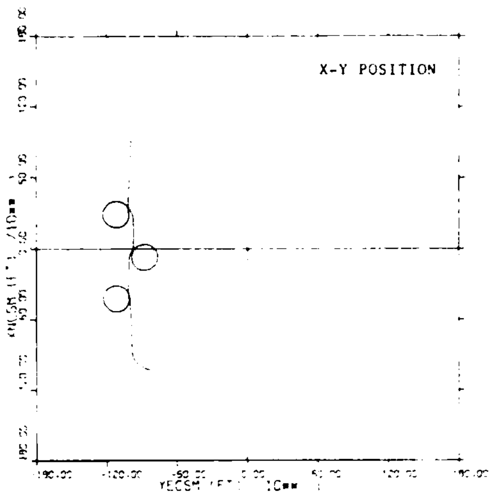


Figure 9. X-Y Position, Depth, And Pitch, Roll And Heading Angle Histories For A Typical CSTV Horizontal Maneuver (3 of 3).



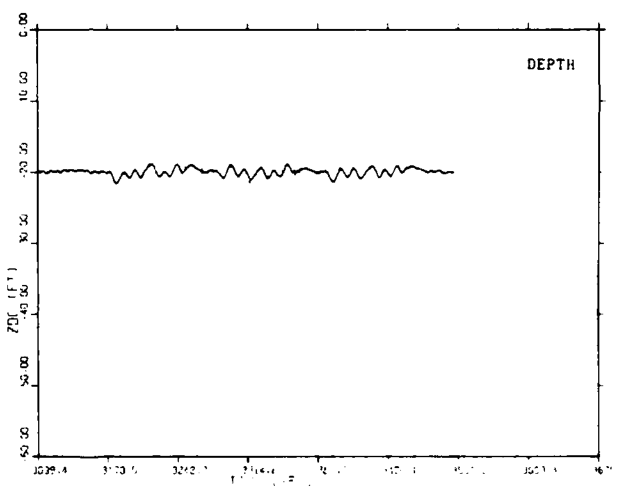


Figure 10. X-Y Position, Depth, And Pitch, Roll, And Heading Angle Histories For A Coupled CSTV Maneuver (1 of 3).

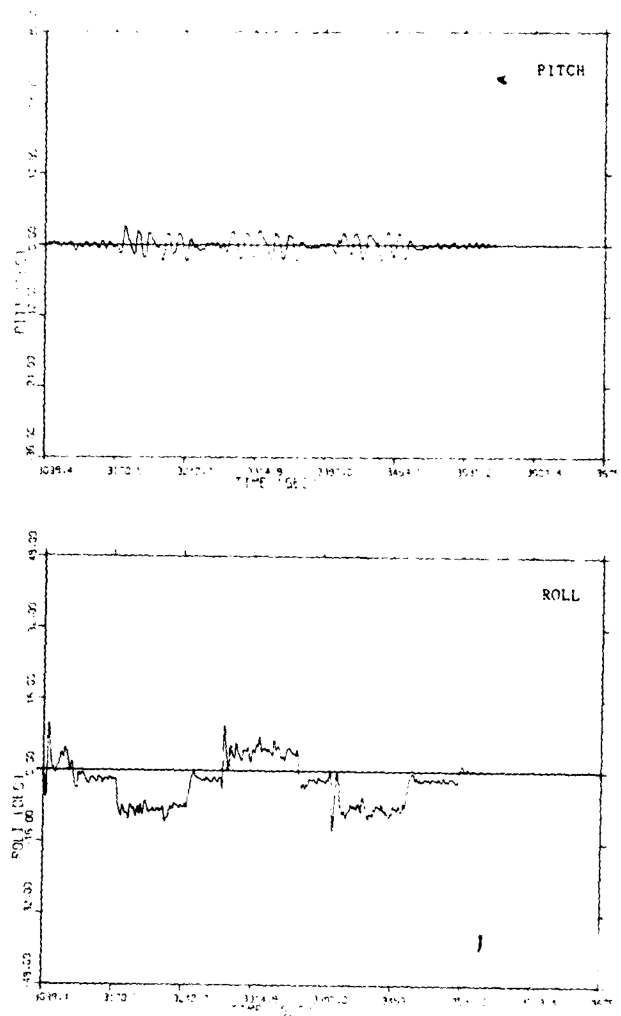


Figure 10. X-Y Position, Depth, And Pitch, Roll, And Heading Angle Histories For A Coupled CSTV Maneuver (2 of 3).

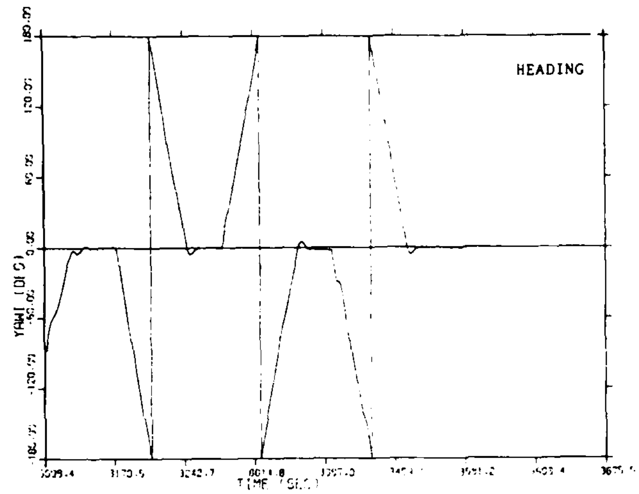


Figure 10. X-Y Position, Depth, And Pitch, Roll, And Heading Angle Histories For A Coupled CSTV Maneuver (3 of 3).

The final set of test data presented in Figure 10 corresponds to a maneuver in which the autopilot is commanded to conduct a series of constant radius turns while commanding pitch excursions to excite cross-coupling dynamics.

CONCLUSIONS

The Control System Test Vehicle has been demonstrated to be a valuable tool for the collection of dynamic response data for submarine configurations. The data collected during the System Identification maneuvers is currently being analyzed by the Naval Sea Systems Command for use in providing new insight into submarine dynamics and maneuverability. To date, only a small subset of the vehicle geometry matrix has been investigated with dynamic maneuvers.

While valuable data have been obtained for the Advanced Submarine Control Program, other vehicle programs at the Naval Coastal Systems Center have been significantly enhanced through the transfer of general autonomous vehicle technology.

REFERENCES

(1) Sejd, J. J., Watkinson, K. W., and Hill, W. F.; "A Submarine Control System Test Vehicle," Naval Engineers Journal, April 1980.

(2) Hall, W. E., and Gupta, N. R.; "System Identification for Non-Linear Aerodynamic Flight Regimes," Volume On Space And Rockets, Volume 14, Number 2 (Feb. 1977).

(3) Gupta, N. R., and Hall, W. E.; "System Identification Technology For Re-Entry Vehicle Aerodynamic Coefficients," Volume On Guidance and Control, Volume 2, Number 2 (Mar. - Apr. 1979).

PAPER WITHDRAWN

PRECEDING PAGE BLANK

DESIGN OF PROPULSION CONTROL INTERFACES FOR EFFECTIVE OPERATOR RESPONSES TO SYSTEM MALFUNCTIONS

by H. L. Williams
and D. B. Malkoff
Navy Personnel Research and Development Center

ABSTRACT

The time has come to reevaluate the roles of the computer and operator in the monitoring and control of malfunctions in gas turbine propulsion systems. Most operators require several years of on-th-job training to become proficient in responding to malfunctions. The computer may be able to do a better job. It can store, retrieve, and process large quantities of information in minimum time. These are essential capabilities for diagnosing the causes of malfunctions and determining appropriate corrective action. The operator will continue to play an important role. What is needed is a better allocation of functions at the control interface, so that full advantage can be taken of the capabilities of both the computer and the operator.

INTRODUCTION

The Seriousness of the Problem

When malfunctions occur in gas turbine propulsion systems, immediate corrective action is often required to avoid damage to equipment, injury to personnel, and/or loss of ship mobility. The malfunctions can result from a variety of causes:

1. Gas turbine engines generate rapid accelerations and rotational speeds; if control over the turbine is lost, the consequences can be catastrophic.

2. The engines are susceptible to foreign object damage. For example, if solid matter, such as frozen particles from a buildup of ice, is allowed to enter the turbine, the turbine blades could be severely damaged.

3. A leak in an oil line or valve would not only create a fire hazard but also pose a threat to the safety of operations personnel because synthetic oils are toxic to humans.

4. Fuel impurities can lead to rapid engine power loss, resulting in large elevations in operating temperature that could lead, in turn, to weakened metal engine structures.

There is every indication that these and other situations can cause malfunctions that lead to unacceptable costs involving lives, missions, and financial outlays.

Current Management Methods

Gas turbine propelled ships in the U.S. Navy employ both computer and human operators to monitor the propulsion system and take corrective action when a malfunction occurs. Computers monitor equipment status, provide status information to the operator, and, in the case of a small number of critical malfunctions, take corrective actions (e.g., shutting down the system). In most instances, human operators diagnose the cause of malfunctions, determine the proper response, and initiate corrective actions.

The Need for Reassessment

Because of the growing complexity of the propulsion system, the increasingly critical nature of its military role, and the possibly devastating consequences of its malfunctions, it is vital that operating personnel make rapid, effective responses. Most operators require several years of on-the-job training to become proficient in this role. For some of the tasks involved, the operator may never be able to become truly proficient, because of the limitations of human capabilities. In those areas, the computer may be able to do the better job. The time has come, therefore, to reevaluate the roles of the computer and the operator in the monitoring and control of malfunctions in gas turbine propulsion systems and incorporate improvements/changes into the next generation of propulsion systems. Undoubtedly, the computer and the operator will both continue to perform important functions; however, a better allocation of those functions is needed. Likewise, a better marriage of computer/operator functions is needed at the control interface so that full advantage can be taken of the capabilities of each.

MONITORING/CONTROL OF MALFUNCTIONS IN CURRENT OPERATIONAL SYSTEMS

Handling Sensory Input

Gas turbine propulsion systems in operational ships are monitored and controlled at a console located in a central control station. Hundreds of sensors installed at strategic locations in the system monitor the status of components by measuring such parameters as temperatures, pressures, rotational speeds, and liquid levels. These measurements are transmitted by means of signal converters and/or the computer to the console where they can be checked by the operator to verify proper system operation. If a parameter falls outside set limits that bound the designated normal operating range, a signal is transmitted to the console. The signal triggers an audio alarm and activates a visual alarm signal to alert the operator that a malfunction has occurred.

Monitoring Inputs and Detecting Malfunctions

In a great many instances, the alarm signals transmitted to the console identify the symptom but not the actual cause of the malfunction. For example, an alarm may alert the operator to a high temperature in a critical bearing. This temperature could be due to a fault in the bearing or to insufficient lubrication oil. The lack of lubrication oil, in turn, could be due to a faulty pump, a clogged filter, or a leak in the oil line. Moreover, a gas turbine propulsion system is made up of many interacting subsystems. An alarm in one subsystem may actually be the symptom of a malfunction in another. The operator, in attempting to diagnose the malfunction, may have to evaluate the performance of a number of interconnected subsystems. He must do this by examining the current and past values of numerous parameters, such as temperature, pressure, speed, and liquid levels, all of which are capable of changing significantly within seconds.

Obviously then, before the operator can be proficient at diagnosing malfunctions, he must be familiar with the behavioral characteristics of all of the many propulsion subsystems and be well-versed in how they relate to one another. He must not only understand the significance of the alarms triggered by hundreds of sensors but also be familiar with the implications of variations within the normal operating ranges of the parameters monitored by these sensors. Lastly, he must be able to relate patterns of measures (i.e., symptoms) from several subsystems before he can diagnose a particular malfunction in one of them.

Decisions in Response to Malfunctions

When a malfunction of a serious nature occurs, the operator seldom has the luxury of time to refer to a manual or wait to obtain additional information from direct visual inspection of remote subsystems. Instead, he must take immediate action based upon what he knows at that moment about the system and its operation. Only a few operators become truly proficient in this area of real-time malfunction detection and diagnosis.

After the operator has diagnosed the cause of a malfunction, he must determine the corrective action to be taken. To do this, he must be familiar with courses of action that are appropriate for any of hundreds of possible malfunctions. Simple solutions, such as shutting down an equipment and/or bringing another equipment on-line, are not always practicable. The operator wants to employ corrective action that will not adversely affect ship performance and interfere with its overall mission. If this cannot be done, he must first notify his superior in the chain-of-command that the equipment needs to be shut down, and receive permission to do so before he can proceed.

CHARACTERISTICS AND CAPABILITIES OF PROPULSION SYSTEM OPERATORS

Sharing of Responsibilities

The propulsion system operator aboard U.S. Navy ships does not work entirely alone when he is diagnosing malfunctions and determining corrective action. An engineering officer-of-the-watch (EOOW), usually a senior petty officer with extensive operational experience, is stationed in the central control station in close proximity to the control console. In the event of a major malfunction, the EOOW will probably become involved in the diagnostic and corrective action processes. Although the participation of a knowledgeable EOOW in these processes can be very beneficial, it can create a highly undesirable situation in other regards. The EOOW is responsible for the supervision of the entire engineering watch team. As such, he may have to cope with multiple, simultaneous emergencies in widely separated spaces in the ship, particularly during combat situations. If he must concentrate his attention on the control console, he cannot give full attention to his other responsibilities. Therefore, it is highly desirable that the operator be trained so that he can handle the system independently.

Formal Training

Propulsion console operators in the U.S. Navy generally are graduates of the Navy's Propulsion Engineering School. Trainees at the school are taught the theory of propulsion system operation and, to a limited extent, are given an opportunity to practice on a highly realistic simulation of the propulsion system that they will operate later on board the ship. Although this training program is an excellent one, it does not result in fully qualified propulsion system operators. Graduates of the school are thoroughly oriented as to propulsion plant theory and operation; however, they are by no means qualified to take over as operators of the propulsion console. The system is much too complex and the knowledge required to handle the job much too great.

On-the-job Training

When graduates of the Propulsion Engineering School report to the ship, they must embark upon a lengthy period of on-the-job training. A number of factors, however, make it difficult for them to become proficient at responding to malfunctions. First, although an operator must be prepared to respond to any one of a large number of possible malfunctions, malfunctions do not occur frequently, particularly major ones. Further, since a ship spends a limited amount of time at sea, opportunities for on-the-job training are limited. Finally, first-tour personnel spend, on the average, less than 2 years aboard ship. At the end of the tour, they may leave the Navy or be assigned to a period of shore duty, further limiting their opportunities to become proficient as console operators.

Continued Development of Skills

Personnel returning for a second tour aboard ship will usually have spent a number of years ashore. Obviously, during this period, their operator skills will have deteriorated and they will have forgotten some of the knowledge required. As a result, many of them will be well into their second tour before they become requalified as console operators.

Decisions regarding corrective actions falling in between these two extremes generally should be assigned to the operator, for a number of reasons. First, in a great many instances, corrective action taken in response to a malfunction affects the ship's general performance. For example, shutting down a gas turbine engine may cause a loss of power. If the ship is in the process of making a critical maneuver, loss of power may have highly undesirable consequences. As a general rule, barring an impending catastrophic event, any action that is going to affect ship performance should be taken only with the knowledge and consent of the commanding officer or the officer of the deck (OOD). The operator must inform these individuals of the need for such corrective action when it arises and obtain their consent before initiating it.

Despite these general guidelines, it must be recognized that no rigid set of operational procedures can or will work to the best advantage in any-and-all situations, many of which are unpredictable. Flexibility must be a characteristic of the operator/computer relationship. A fully automatic diagnosis/control mode option should be available, should that be necessary for ship survival. Likewise, manual override options should be available for those instances where the automation fails or contradicts the mission of the ship.

Other Views of Human and Computer Roles

Some persons feel that computers will never be able to replace the human element at higher levels of decision making, and, therefore, all real-time control decisions must be human in origin. However, recent developments in the field of artificial intelligence and, in particular, the area of expert systems, demonstrate that a computer program can act effectively in a manner that performs at least as well as the very best of human experts. Indeed, research continues to explore ways in which such programs might be able to learn from experience so as to outperform their human counterparts.

Others feel that, for social or moral reasons, regardless of (and perhaps because of) the remarkable capabilities of computers, control automation must never be permitted, even at the most rudimentary levels. This attitude conflicts with the current situation, where many of our needs simply cannot be met without the use of automation. Certainly, this is true in the field of propulsion engineering, where little social or moral comfort can be derived from an accident that wreaks havoc aboard ship.

The real challenge, then, is how to amalgamate the two--humans and computers--so as to take advantage of the best abilities of both, while avoiding the pitfalls of either. It seems that the solution, at least in part, would be the careful, wise allocation of man-machine functions between the operator and the computer.

DESIGN FEATURES OF THE PROPULSION CONTROL INTERFACE

Implementation Implications of the Concept of Divided Functions

If functions are to be divided between operator and computer, it is imperative that the control interface be designed to facilitate the joint operation. If the operator must make important decisions relative to implementing corrective action recommended by the computer, he must be provided with the necessary supporting information to enable him to do so. He must, as stated earlier, be made aware of any malfunction immediately following its occurrence. Likewise, if the computer is aware of an impending malfunction, it should inform the operator immediately.

Both audio and visual alarms should be used to alert the operator; and the CRT or plasma display computer screen, to provide whatever information is available about the nature of the alarm (e.g., identification of the subsystem containing the parameter that has exceeded normal limits). The amount of available information concerning the nature of the malfunction causing an alarm often exceeds that which can be conveyed by the label for an alarm indicator light.

Next, the computer should present, via a CRT or plasma display screen, its diagnosis of the cause of the malfunction. Enough information should be presented to allow the relatively inexperienced operator to comprehend what it is all about. The information should be presented in a summary form, however, that allows the experienced operator to comprehend the nature of the malfunction without having to wade through details that he does not need.

Finally, the computer should present recommended corrective action to the operator. The same logic should be followed as outlined above for presentation of the diagnosis. A reasonable amount of detail should be presented in a form that can be assimilated by the experienced operator at a glance. Additional, more detailed information should be available for the operator should he request it. The computer program must be able to provide this in the sequence selected by the operator, at the time he requires it, in a manner that is clear and allows the option to terminate the sequence at any time.

If a malfunction occurs for which the computer has no diagnosis, it should so inform the operator. If it can do so, it should recommend the steps the operator should take to prevent damage to the equipment and list the specific additional information that would enable the diagnosis to be made.

Expert Systems as a Tool for Implementation

Obviously, a computer program that will do the right thing at the right time, in the most efficient and clear manner, must embody within it a considerable amount of wisdom. Programs of this sort that contain large amounts of knowledge based upon and obtained from subject matter experts are referred to as "expert systems." Such systems often use rules of thumb learned by human experts as the result of many years of experience in the field. An expert system designed expressly for malfunction detection and propulsion control could offer enormous advantages. It could provide the very best of advice to all propulsion control operators, regardless of their own personal limitations in training or experience. That advice could be made available instantaneously, in a form easily understood. Systems of this kind could be constructed with various degrees of automation or manual control, or both. They could even have the capability to explain the reasons for their decisions, ask for missing data, and learn from experience. In a gas turbine propulsion environment, where little is known of all the subtle complex patterns of sensory input that might indicate an early trend toward disaster, these learning modes can conceivably outperform their expert designers.

RECOMMENDED PROGRAM OF STUDY

Design of the propulsion control interface described above will require an in-depth systems analysis that goes beyond anything the Navy has heretofore attempted at such an early stage of development. Prior to design of the control interface, all expected critical and frequently occurring malfunctions must be identified. Sensors must be identified and their outputs described. Symptoms of the malfunctions, as they are likely to appear at the control interface, must be identified and the diagnostic process defined. Finally, diagnostic outcomes must be related to recommended corrective action.

Considering the complexity of the gas turbine propulsion system in a modern Navy combatant, this task may appear on the surface to be all but impossible. However, if we consider such a task impossible for the sophisticated propulsion engineer and computer scientist, then it most certainly is impossible if left solely to the console operator aboard ship. On the contrary, we feel that great strides are possible in this area. Accordingly, the Navy Personnel Research and Development Center has recommended a program of study to define the characteristics of an improved control system. Steps in this program are as follows:

1. Determine the required content and best format of alarm messages on the CRT or plasma display.

2. Define the characteristics of the computerized diagnostic processes, including required sensor information and computer processing functions.

3. Determine the required content and best format of diagnostic messages to the operators.

4. Identify the types of corrective action the computer is likely to recommend and determine the required content and format of messages to relay this information to the operators.

5. Identify the types of corrective action that should be initiated by the computer and by the operator.

6. Determine how to exploit fully the capabilities of the computer and apply this knowledge in the design of the propulsion control system.

The time to get started on this program is now, prior to the development of next-generation propulsion systems. Successful completion of the steps listed will allow us to fully exploit the capabilities of the computer in the control and monitoring of propulison systems. As a result, the computer can be used to diagnose the causes of malfunctions and recommend appropriate corrective action, and the operator, when time permits, can take the final step in initiating the corrective action.

A MODERN MACHINERY CONTROL CONSOLE FOR WARSHIPS

LCdr P.J. MacGillivray and LCdr D.J. Marshall
Canadian Forces (Navy)
Department of National Defence (CANADA)

ABSTRACT

The Canadian Navy has developed an ergonomic machinery control console, incorporating colour CRT displays and computer-generated graphics as the sole vehicle for information display. The selective presentation of information enables the watchkeeper to interact with the machinery systems in a more efficient manner than has previously been achieved.

This console has been designated as the Standard Machinery Control Console (SMCC). In a militarized form, it can serve as the main machinery console for any class of modern warship. This paper describes the origins and objectives of the project and presents the salient features of the SMCC in its present form.

INTRODUCTION

The Standard Machinery Control Console (SMCC) project was established as a parallel development to the SHipboard INtegrated MAchinery Control System (SHINMACS) program. SHINMACS provides for fully automatic control and surveillance of all shipboard propulsion, ancillary, auxiliary and electrical-generating machinery through a distributed, digital processor-based control system. Previous meetings of this symposium have witnessed the evolution of this concept (1,2,3,4); its adaptation in the Canadian Patrol Frigate is reported in (5).

Much of the Canadian Navy's recent work has centred around the man-machine interface (MMI) itself. The ergonomic requirements for this unique MMI were defined by the Defence and Civil Institute for Environmental Medicine (DCIEM) (6,7,8). Parallel research and development activities produced a real-time simulation of the DDH-280 propulsion plant, and a prototype MMI. The integration of these two products is the SMCC itself, pictured in Figure 1.

Figure 1 - The Standard Machinery Control Console

The objective of the SMCC project was to design, develop, and manufacture a prototype of the SHINMACS console (later defined as the SMCC) which would be used to:

a. demonstrate and validate the concepts of the ergonomically-designed console;

b. develop, assess and, if necessary, modify the colour graphics presented to the watchkeeper; and

c. assist in the development of training requirements for watch-keepers who would interact with SHINMACS.

Having established these goals, the project embraced the following *constraints and principles:*

a. strict adherence to the ergonomic aspects of the console design;

b. use of commercial grade hardware to keep the required funding within the limits of minor R & D projects;

c. use of high resolution colour displays to properly simulate the final (militarized) system characteristics; and

d. provision for easy refinement of the information display structure.

SMCC SYSTEM DESCRIPTION

The major components of the system are as described in the following paragraphs and as depicted in Figure 2.

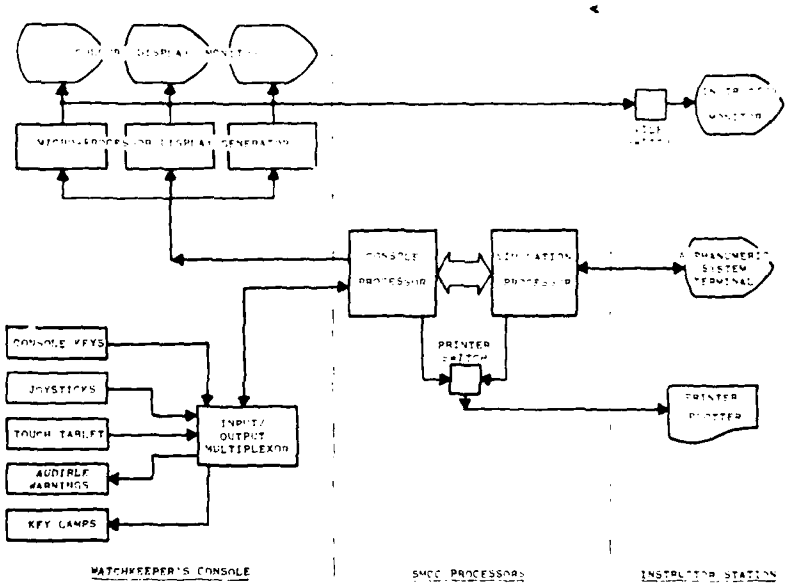


Figure 2 - SMCC Functional Structure

The Machinery Control Console is a one-man workstation housing three high resolution colour CRT displays, the control panels, and a work surface. Operator inputs consist of force joysticks, push-button keys, and a touch tablet.

The Console Processor interprets control panel commands, passes the information to the simulation processor and generates the graphics pages. Static information is retrieved from disk, displayed, and overlaid with dynamic values obtained from the simulation processor.

The Simulation Processor reacts to inputs from the console processor and the instructor's station, and executes a real-time simulation of the DDH-280 main propulsion machinery and a rudimentary simulation of the associated ancillary systems. Approximately 300 sensor points are simulated.

The Instructor's Station consists of one high resolution colour CRT and a monochrome alphanumeric terminal. The supervisor is able to alter plant status, simulate propulsion plant control from the bridge, monitor the watchkeeper's screens, and execute graphic print routines.

THE MAN-MACHINE INTERFACE

All machinery plant information is displayed using computer-generated graphics. Information is presented in a meaningful format, both spatially and temporally, such that the information required for each operator task is readily available and does not induce adverse levels of operator workload. The static portion of each page is augmented by dynamic information, consisting mainly of discrete digital parameters, bar graphs, and graphical status of individual components such as engines, pumps, and valves. There are presently 42 pages of information available to the watchkeeper, excluding up to three pages for storage of additional alarm and warning information.

Primary information is accessed via dedicated keys grouped in the following modules:

a. Overview Pages Module.

(1) Propulsion Overview - (see Figure 3) This page consists of: active mimics of each propulsion train; digital and analogue presentation of the four major parameters for each driving engine; discrete values of shaft power, torque, RPM, and propeller pitch; a telegraph module displaying ordered, reply, and actual power commands; four control blocks which access alternate pitch/RPM schedules and control modes, transfer Station-in-Control, and allow operator control of offline engines. This page is automatically displayed following changes to demanded power levels or requests for a Station-in-Control transfer.

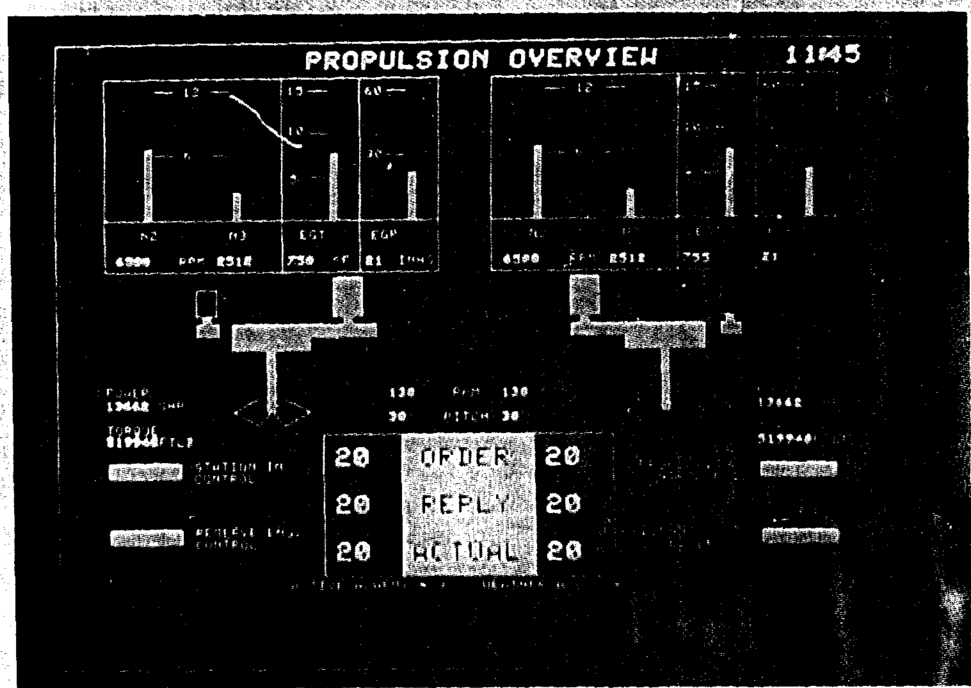


Figure 3 - Propulsion Overview

[illegible] provides an indication of the status of every system.

(3) Alarms Overview - see Figure 4. Up to 72 alarms and warnings are stored and displayed on a total of four pages in a time-ordered sequence. This page is displayed automatically with each new alarm or warning and following the clearing of any previous abnormal condition. Each entry consists of: point status (alarm, warning, or cleared); system and sensor names; fault type; low and high limits (if applicable); present value; point identification; time of entry.

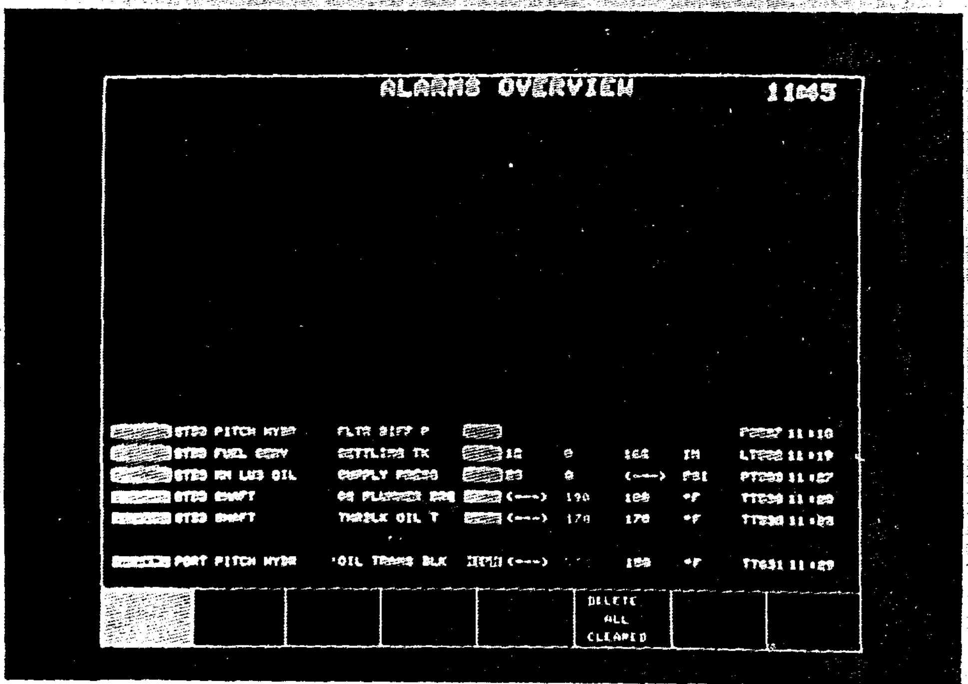


Figure 4 - Alarms Overview

(4) Shaft Data (2) - One page is provided for each shaft and consists of: clutch, gearbox and shaft-line mimics; scan point indication of shaft-line sensors, together with sensor information; control block enabling remote operation of the cruise engine clutch locking arrangement.

(5) Operator Monitor - (see Figure 5). The operator may select up to 12 system parameters and graphically display the rolling [illegible] minute history of any of these 12 points. Each log graph consists of: actual and relative time; current parameter value; alarm and/or warning levels as applicable; full scan point identification.

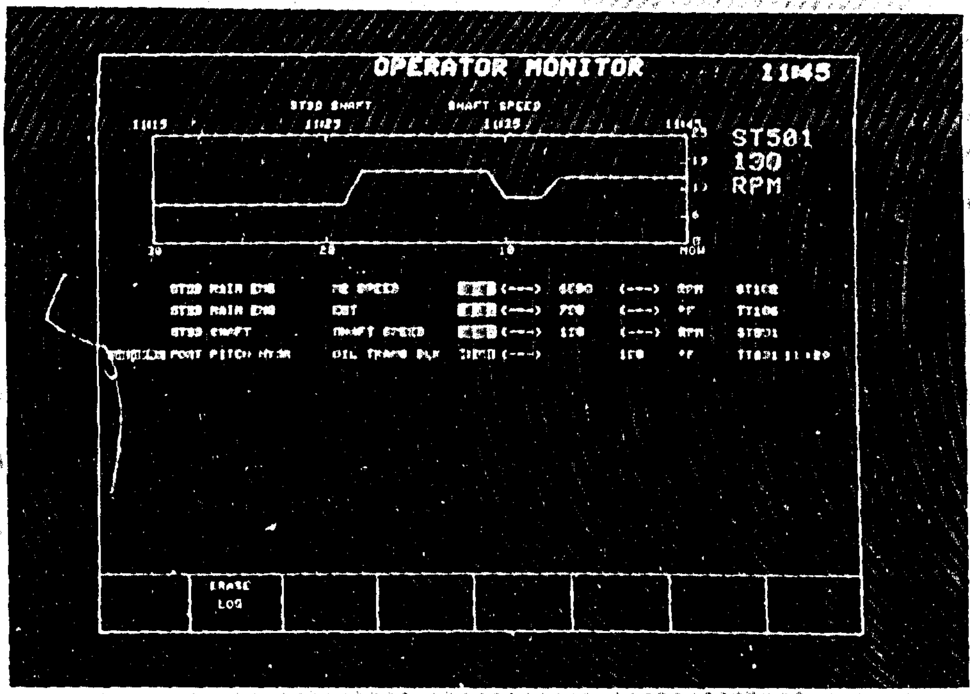


Figure 5 - Operator Monitor

b. Engine Pages Module.

(1) Engine Control - (see Figure 6) One page is dedicated to each gas turbine and consists of: a digital and analogue display of the four major parameters for that engine; a memo pad which logs significant events occurring in a complete cycle. The operator may start, stop and assume power via the automatic sequences or may manually initiate the start and stop sequences.

(2) Engine Data - (see Figure 7) One page is dedicated to each gas turbine consisting of: a mimic of the gas generator, free power turbine and turbine brake; scan point indication of engine sensors and sensor information.

(3) Engine Log - One page is dedicated to each engine set and, in a manner similar to the Operator Monitor, simultaneously graphs the 30 minute rolling history of the four major engine parameters.

(4) Driving Engines Log - This page is a composite of the four separate Engine Logs, enabling the operator to compare the 30 minute rolling history and, therefore, the performance of similar engines.

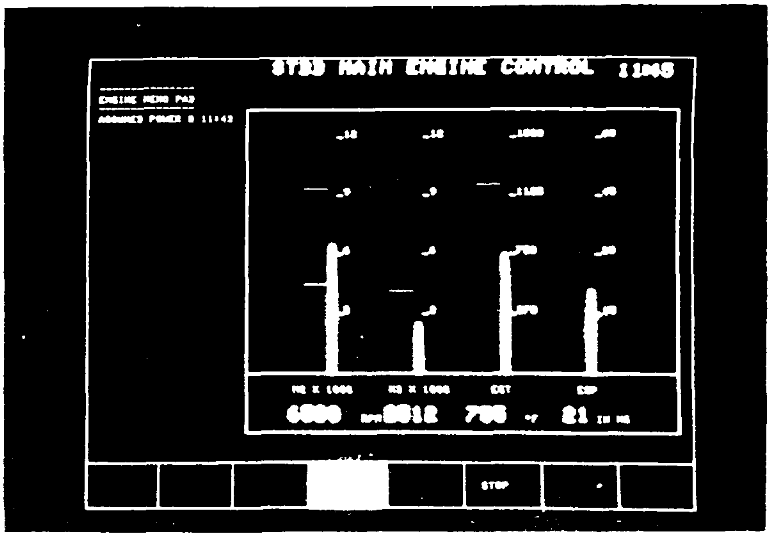
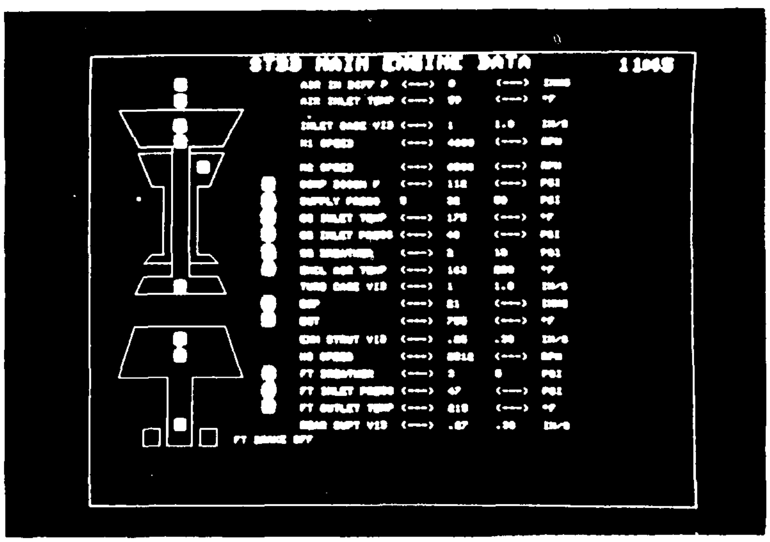


Figure   Engine Control Pa



The secondary information, consisting mainly of active process flow diagrams, are accessed primarily through dedicated keys grouped in the following modules:

a. Ancillary Systems Module.

Main Lubricating Oil System
Engine Fuel Service System
(Engine) Synthetic Lubricating Oil System
Propeller Pitch Hydraulic System
Engine Hydraulic Start System
Sea Water Cooling System
Engine Anti-Icing System
Main Reduction Gearing (2)
Engine Fire Detection and Extinguishing System
Engine Water Wash System
Fuel Purifier System

Figure 9 illustrates the Propeller Pitch Hydraulic System, just one of the sixteen active process flow pages available.

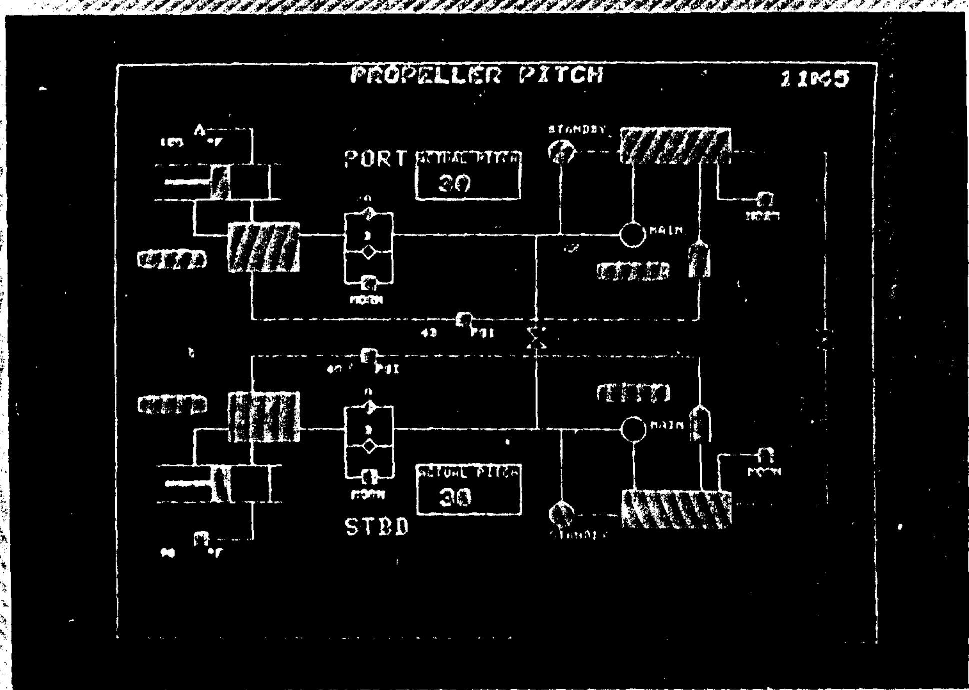


Figure 9 - Propeller Pitch Hydraulic System

[illegible]

- Air Conditioning Plants
- Domestic Steam System
- Main Refrigeration System
- Fresh Water System
- Steering System
- Fire Pumps and Firemain System
- HP and LP Air Systems

Within the SMCC, none of these systems are actively simulated; nevertheless, Figure 10 indicates the bases of these displays which, in a dynamic form, will be expanded upon in fully developed systems.

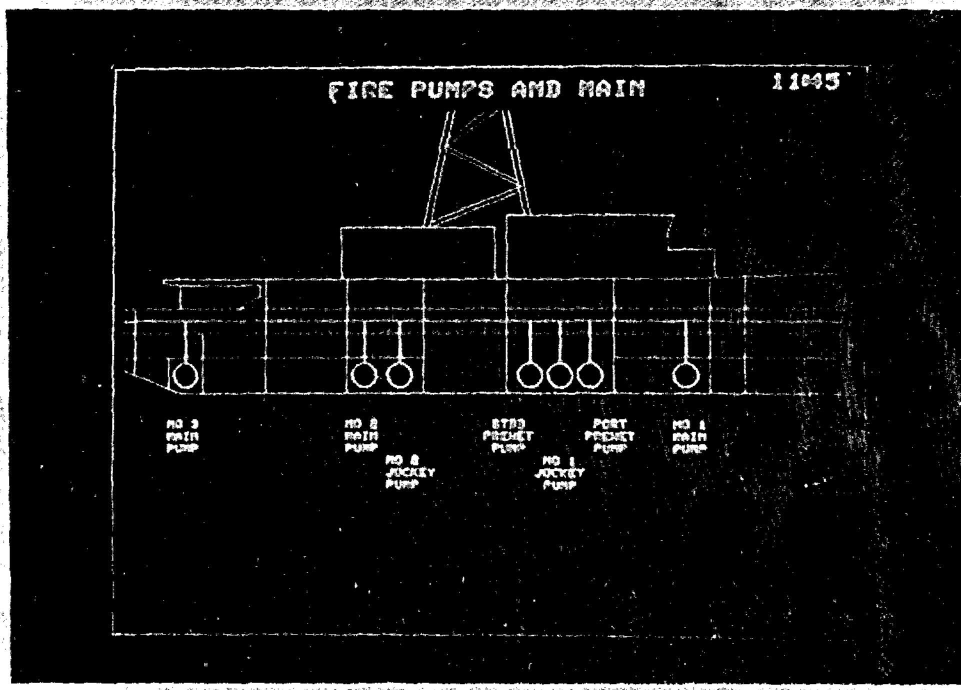


Figure 10 - Fire Pumps and Firemain System

The operator may direct the display of information on any particular screen. Alternately, he may simply select the desired [illegible] or opt to have the console display the information on the [illegible] screen on a first-in/first-out basis. The watchkeeper is also given the option of protecting up to two displayed pages.

The operator inputs to the system [illegible] dedicated push button keys, a touch tablet, [illegible] sets of [illegible] dedicated (soft) line menu option keys located below each [illegible] screen.

A separate module of eight dedicated [illegible] to enable the engines for starting and, if necessary, [illegible] once the starting sequence has been initiated.

In the automatic propulsion mode, two force joysticks enable the operator to set the shaft power level according to one of a limited number of defined shaft RPM/propeller pitch schedules. Under certain circumstances, he may select a link feature and set both shaft power levels using either joystick. In the semi-automatic propulsion mode, he has independent control of engine power and propeller pitch.

Specific commands relating to pump and valve operation are input to the plant by means of the touch tablet and line menu option keys. The touch tablet is used to position the cursor over the control block associated with a particular component. Selection of that control block generates a line menu of options. Finally, depression of the associated soft key causes the comand to be executed. It should be noted that only the valid options associated with a particular component are generated. For example, the watchkeeper is not permitted to start a particular pump if its associated suction valve is not open.

EVALUATION OF OPERATOR INTERACTION

The primary objective of the SMCC project was to demonstrate and validate the concepts of the ergonomic design specification. In order to validate the assertion that the SMCC could form the basis for an effective MMI, DCIEM was tasked to conduct a human engineering evaluation of the console. The evaluation process, results, and recommendations are fully reported in (9,10).

The evaluation was constrained for the following two reasons:

a. The prototype SMCC was never intended to be used as a trainer or real-time simulator but was built only for demonstration purposes. Consequently, funding was not requested for incorporation of data-logging facilities.

b. The follow-on SHINMACS Advanced Development Model (ADM) had been approved and was underway when the evaluation commenced. Recommendations for changes in hardware configuration could only be incorporated in the ADM if tendered quickly.

Despite these limitations, a total of 43 hardware and software changes were recommended. Virtually all of these have been incorporated in the next stage of development. In view of the selection of SHINMACS as the machinery control system for the CPF (6), the results of these evaluation efforts will also be seen in the final production version and at sea. With respect to the more significant problems addressed in (10), the following comments are provided:

a. The problem of possible display-control incompatibility has been addressed by drawing the page title first and flashing the title in amber should an incompatibility condition arise. A recommendation to automatically re-arrange pages has been rejected. Under damage conditions, such re-arrangement could impair the operator's ability to display needed information.

b. The joystick reply display update rate has been improved in the SMCC through modifications to the console software. The CPF specification contains a much tighter requirement.

c. Although not directly related to the SMCC, a detailed evaluation of the SHINMACS ADM is to be conducted in the fall of 1985. Data-logging facilities will be available.

CONCLUSION

The Standard Machinery Control Console has served the Canadian Navy's purposes particularly well. Following a comprehensive definition of the ergonomic requirements, and for a relatively modest outlay of R & D funds, the prime objectives were satisfied. It has supported the proposition that a colour CRT-based man machine interface is not only viable, but is likely preferable to conventional machinery control consoles. The SMCC is inherently flexible and is not tied to any particular class of ship. Consequently, it forms the basis for a machinery control console which will likely see service in all classes of Canadian warships.

REFERENCES

(1) Gorrell, E.L., "Human Engineering Considerations In Design of Information Displays For Machinery Control of Future Naval Ships", Proc. FIFTH Ship Control Systems Symposium, Vol 4, David W. Taylor Naval Research and Development Center, Annapolis, MD, USA, 1978.

(2) Ironside, Cdr J.E., "SHINPADS An Integration Philosophy For The 21st Century", Proc. SIXTH Ship Control Systems Symposium, Vol 1, National Defence Headquarters, Ottawa, Canada, 1981.

(3) Gorrell, E.L., "SHINMACS Machinery Control Console Design", Proc. SIXTH Ship Control Systems Symposium, Vol 4, National Defence Headquarters, Ottawa, Canada, 1981.

(4) Baxter, Cdr B.H., LCdr R.J. Rhodenizer, and P.V. Penny, "Shipboard Integrated Machinery Control System (SHINMACS) - A Canadian Forces Concept", Proc. SIXTH Ship Control Systems Symposium, Vol 5, National Defence Headquarters, Ottawa, Canada, 1981.

(5) Eich, H.P., and R. Khan, "The Machinery Control System For The Canadian Patrol Frigate", Proc. SEVENTH Ship Control Systems Symposium, Bath, England, 1984.

(6) Gorrell, E.L., "Human Engineering Design Requirements For SHINMACS Machinery Control Consoles: Part II. Console Hardware and Functions", DCIEM Technical Report No. 80-R-27 (For Defence Purposes Only), Downsview, Ontario, Canada, 1980.

(7) Gorrell, E.L., "Human Engineering Design Requirements For SHINMACS Machinery Control Consoles: Part III. Console Displays and Operation", DCIEM Technical Report No. 81-R-18, Downsview, Ontario, Canada, 1981.

(8) Gorrell, E.L., "SHINMACS Machinery Control Console Design", DCIEM Report No. 81-R-27, Downsview, Ontario, Canada, 1981.

(9) Gaines, B.R. and N. Moray, "Human Engineering Evaluation of Standard Machinery Control Console System", SHINMACS-HEE-FIN:MAR 84.

(10) Turner, D.C. and D. Beevis, "Human Engineering Evaluation Of A Digital Machinery Control Console (MCC): A Case Study", Proc. SEVENTH Ship Control Systems Symposium, Bath, England, 1984.

HUMAN ENGINEERING EVALUATION OF A DIGITAL MACHINERY CONTROL CONSOLE (MCC): A CASE STUDY

D.C. Turner and D. Beevis
Defence and Civil Institute of Environmental Medicine
Toronto, Ontario, Canada

ABSTRACT

As digital machinery control systems move from the development phase into the implementation phase, it is necessary to verify the concepts underlying the replacement of conventional process control instrumentation with computer graphic information displays.

This paper presents a case study of a human engineering evaluation of a watchkeeper's machinery control console for a digital control system. Results concerning the impact of delays in display updating, display-control compatibility, display topography, and operator/console model matching are discussed. Recommendations stemming from the results are made as well as requirements for future evaluations of similar systems.

INTRODUCTION

The development of a digital machinery control system called Shipboard Integrated Machinery Control System (SHINMACS) and its human factors considerations have been reported at the FIFTH and SIXTH meetings of this symposium (1,2, 3,4,5). The watchkeeper's MCC discussed in this paper is a demonstration model of the SHINMACS watchkeeper's console. The implementation of digital systems, such as SHINMACS, provides unique capabilities for displaying information compared with the capabilities of conventional process control instrumentation. (Conventional systems can be defined as information display by means of vertical and circular scale analogue meters, digital panel meters, annunciator matrices, and teletype writers.)

Several inherent disadvantages of conventional systems have been noted by Gorrell (2). The size of conventional panels requires excessive visual scanning to obtain a comprehensive pattern of displayed information and hence machinery status. Size also limits the application of mimic diagrams to reinforce the watchkeeper's internal model of the machinery control system. Attempts to condense the information display by the use of digital panel meters and teleprinters can place unnecessary memory loads on the watchkeeper.

Digital systems such as SHIMMACS can exploit the flexibility of electro-optical displays by presenting information in task related chunks. Within chunks, the information can be formatted in process flow and mimic diagrams. It has been argued that the use of such diagrams can reinforce the watchkeeper's internal model of the machinery control system, thereby reducing operator memory and cognitive workload (2,4).

The purpose of this paper is to summarize the major results of the human engineering evaluation of the watchkeeper's MCC (6). The evaluation was conducted to verify the concepts for the man-machine interface which had previously been developed (7,8,9). The primary concept was that CRT screens can be used to display information in a coherent and meaningful format, both spatially and temporally. Another important concept was that the information required for each operator task can be made readily available without inducing adverse levels of operator workload. Equally important was the assumption that the temporal sampling of information through CRT screens, rather than conventional process control instrumentation, would be an acceptable and adequate mode of operation.

BACKGROUND

In support of SHINMACS development, the Canadian Forces Directorate of Marine and Electrical Engineering (DMEE) constructed a system called the Standard Machinery Control Console (SMCC) System to demonstrate the SHINMACS concept (10, 11). The SMCC System models the operation and control of a DDH 280 class destroyer's main propulsion system, ancillary systems, and selected auxiliary systems. The purpose of the SMCC System was to demonstrate and evaluate the information displays and controls of the watchkeeper's MCC.

The SMCC System (10) is composed of four major components:

- a Machinery Control Console (MCC),
- a MCC Control Computer,
- a Ship's Plant Simulation Computer,
- an Instructor's Console.

The MCC has three CRT screens, displaying computer-generated, colour process flow and mimic diagrams. Controls consist of push-button keys, joysticks, a touch tablet, and a voice communications module. The keys control information displayed on the screens and some engine functions. The joysticks control engine power and propeller pitch.

The MCC Control Computer executes a program that responds to control inputs, and passes this information to the Simulation Computer. The program also receives information from the Simulation Computer and changes information displayed on the screens as a result of the data received. The Simulation Computer executes a program that simulates the operation of the ship's engines and associated propulsion system components. It produces information that would normally be generated by sensors monitoring equipment from within the ship.

The Instructor's Console provides the ability to feed inputs into the plant Simulation Computer and to monitor the watchkeeper at the MCC.

In November 1983, DMEE tasked DCIEM to let and supervise a research contract for the conduct of a human engineering evaluation of the watchkeeper's MCC. Due to the lead times associated with follow-on contracts, DMEE and DCIEM had just over three months to complete the evaluation, from delivery of the MCC. DMEE agreed to supply and train fleet watchkeepers, and develop evaluation scenarios based on the requirements of the contractors. The contract was let to Man-Machine Systems Consultants Incorporated, and the Centre for Person-Computer Studies Incorporated.

METHOD

At the end of January, 1984, a meeting was held to familiarize the contractors with the MCC and to define an evaluation plan. The SMCC System operator test procedure scenarios used during system acceptance trials were demonstrated by DMEE staff. The contractors examined a range of operational procedures, studied the control and display ergonomics in detail, and operated the MCC themselves.

Following this familiarization, it was agreed that the contractors would assess the system based on their experience and knowledge of process control systems and man-machine interfaces, due to time constraints. They were to provide expert assessment of human factors problems as well as an appraisal of opinions expressed by the watchkeepers after operating the MCC.

The decision to conduct the evaluation using an experiental approach, rather than an experimental design format was necessary because of timescale and other constraints listed below:

1) the short time period available for the study;

2) the lack of proven operational procedures and training manuals;

3) the limited time available for training and running operators;

4) the design of the MCC as a demonstration prototype not an interactive simulator;

5) the lack of electronic data-logging facilities in the SMCC System, which made data collection extremely difficult;

6) the impossibility of rectifying any ergonomics problems on a step-by-step basis, to ensure that other problems had not been masked; and

7) the incomplete documentation of previous human engineering design trade-off decisions.

Given the above constraints, the method described below was developed to allow maximum opportunity to detect human factors problems.

A six-day evaluation plan was developed. A group of four operators with DDH 280 machinery control experience were to be trained over a period of four days. On the last two days of training, one contractor would assess operator training and administer a questionnaire based on observations made during the familiarization visit. On two subsequent days a contractor would assess operator behaviour based on two evaluation scenarios and re-administer the questionniare. Each operator would have a half-day to complete the first scenario followed by the second scenario.

The four operators recruited by DMEE were experienced in DDH 280 machinery control and two were experienced watch supervisors. The first day of training consisted of system familiarization. DMEE staff briefed the operators on the SMCC System concept and reviewed the functions of the MCC displays and controls. Operator manuals (10) were distributed and any questions arising from a personal review were answered. On the second day, three operators were guided through the SMCC System operator test procedure scenarios. On the third day, the fourth operator was guided through the same scenarios. Each operator was also trained on warning/alarm drills. Except for administering the questionnaires, all operators were present during individual training. While one operator practiced at the MCC, the others were observing. As great a range of operational conditions and warning/alarm states were practiced in the time available. A contractor observed training procedures and distributed a questionnaire at the end of the third day. Each operator responded individually in their own time and without discussion. The contractor then debriefed each operator concerning their responses. On the fourth day, each operator repeated the training scenarios without guidance.

During the training period, DMEE co-ordinated the generation of the training and evaluation scenarios. After reviewing the content of the training scenarios, DMEE prepared the two evaluation scenarios. Care was take[illegible] that all the necessary procedures needed to complete the evaluation scenarios had been taught during the training period. DMEE also ran both scenarios to ensure their correctness and to estimate the time necessary for their completion.

The first scenario, approximately 60 minutes long, was designed to take the operator through a range of normal operational procedures. These included speed changes, transfer of station-in-control, locking clutches, washing an engine, manual control, and responding to an engine trip and vibration alarm. The scenario tested the operator's familiarity with the console under routine but demanding, conditions.

The second scenario was approximately 20 minutes long. It was designed, to the requirements of the human factors contractors, to put the operator under the stress of an abnormal condition of "battle damage". The purpose of the second scenario was "to evaluate whether the operator had an adequate conceptual model of the console and its operation" in order to "transfer his experience and devise proper operational procedures for the abnormal conditions" (6). Operational procedures included speed changes, assuming power on other engines, an exhaust gas temperature alarm, a gearbox bearing alarm, a request for gearbox bearing history, an engine fire, transfers between manual and automatic propulsion modes, and a differential pressure alarm. Battle damage was simulated by having only one CRT screen functioning rather than the normal three. Operators had not met this situation in training. The scenario was significant because, to quote the human factors contractors, it "forced the 'information-sampling' mode with the console to be used at its limit. If the operator is successful in the single CRT screen mode then it is reasonable to suppose that the sampling of information through the screens is an acceptable and adequate mode of operation" (6).

The evaluation trials were conducted as planned over a two-day period. The evaluation runs were observed by DMEE 7, DCIEM and a contractor to ensure that specialists in the aspects of SMCC System capability, design concept, and scenario results were available to interpret results on-site. After each evaluation run the observers debriefed as a group to ensure all events occurring during the run were commonly understood. All runs were recorded on video so that any unforeseen events which might have occurred could be reviewed if required.

RESULTS

The SMCC System was, overall, found to be a viable man-machine interface for ship machinery control. The control panel and paging system appeared to be designed appropriately. The design of the graphics display software seemed adequate. Based on operator experience and evaluation results, it was concluded that the system is very natural to use and requires little training for a watchkeeper familiar with the machinery control aspects of a DDH 280 class ship. Operators detected and diagnosed all faults quickly. Although there were some delays and errors in selecting CRT pages for display, these were attributable to a lack of training and the slow response of the demonstration model, rather than to major design faults in the MCC. Even when only one CRT screen was operational, operator performance was remarkably efficient. There seemed to be little difference in their response times compared with when three screens were operating. This finding is attributed to the manner in which the MCC reinforces the watchkeeper's internal model of the ship's machinery control systems.

One major problem detected was the delay in display updating. The response speed between inputs from the control joysticks and the corresponding updating of displays was not adequate. The delay was significant enough to impair operator: system interaction, and hence ship control. The problem made proper appraisal of the pressure sensitive joysticks impossible and may have obscured the detection of her possible problems.

A significant problem identified during the evaluation was the possibility of confusing the left-right/port-starboard stereotype in maintaining compatible display-control relationships. It is a fundamental principle of human engineering that such compatibility be maintained.

Another related problem was based on the importance of ensuring that the orientation or topography of all mimic diagrams and process flow charts is compatible. Two instances of incompatibility were identified.

Finally, some circumstances were discovered where the model presented by the SMCC System confused some of the operators. In general, the operators came to assume that, if a problem existed and was displayed by the MCC, there must be some way of rectifying it from the console without sending a roundsman to complete an action. Operators also assumed that all possible machinery control information would be displayed. This was not and would not always be the case.

DISCUSSION

The type of evaluation performed was very limited in relation to the complexity of the MCC. However, several important limitations of the current design were identified.

The delay found in display updating is attributable to the fact that the SMCC System was intended as a demonstration model not a real-time simulator. The delay is caused by the computers which drive the demonstration model, and the software, which samples every point equally prior to updating the displays. It is intended that, early in the next development phase, this weakness will be re-assessed using upgraded hardware and software.

The problem of not maintaining the left-right/port-starboard stereotype in display-control compatibility stems from the inherent flexibility built into the MCC for page assignment to any of the three CRT screens, and the bias of the console design towards the stereotype. The layout of the controls on the console are such that port and starboard controls are to the left and right of the centre-line respectively. The operator can either have pages assigned automatically by the computer, or manually assign pages to any of the CRT screens. If in the manual mode, the computer still automatically assigns the display of the Alarms Overview Page. The flexibility to assign any page to any screen is useful at times but it is assumed that under normal circumstances the operator will display the Propulsion Overview Page, which summarizes both Port and Starboard propulsion system status, on the centre screen, with Port associated systems displayed on the left screen and Starboard associated systems on the right screen. This complies with the spatial stereotype. When the stereotype is not enforced confusion can arise, as the following incident demonstrated.

The Propulsion Overview Page was displayed on the centre screen. The Starboard Engine Control Page was displayed on the right screen. The left screen was also occupied. A vibration alarm on the Starboard engine was detected by the system and the computer overwrote the right screen with the Alarms Overview Page. The operator responded by displaying the Starboard Main Engine Data Page on the left screen. It is here that the left-right Port-Starboard stereotype was violated. The operator proceeded to reduce Starboard engine speed manually but then became confused. An analysis indicated that the operator manipulated the correct manual starboard control but received incorrect feedback because his attention was drawn to the parameters displayed for the Port propulsion system on the Propulsion Overview Page not the Starboard propulsion system parameters. Naturally the operator could not deduce why the parameters were not responding to his control movements. The explanation (6) given for this was that "the 'left-ness' of the left-hand screen was so powerfully compatible with the 'left-ness' of the *Port engine* mimic that it suppressed the operator from remembering that what was displayed on the left screen was a Starboard engine".

This problem cannot be completely avoided through training and procedures. The restriction of Port displays to the left screen and Starboard displays to the right screen is not an acceptable solution. However, several solutions have been recommended for further evaluation. One idea is for the title of each page to be drawn first rather than last as it is currently. A second idea is that if a Port (Starboard) page is sent to the Starboard (Port) screen, the title should be coloured amber and the title should flash continually. It has also been recommended that the page presentation logic be changed so as to automatically re-arrange pages in an order that has no reversals. For example, any Port pages will be displayed in the left-most positions available and any Starboard pages will be displayed in the right-most positions available.

The problem of topographical incompatibility is important. Because of its importance all but two of the mimic diagrams and process flow charts are drawn with the "ship's head up". It is assumed that the console will be located facing forward. Hence screen topography will be compatible with ship topography.

The exceptions are the mimic diagrams for Port and Starboard Main Reduction Gearing. These are drawn with the ship's head to the right of the screen because of the size of the system. Except for the page titles, the pages are identical. In order to alleviate the topography problem it has been recommended that an outline of the ship's centreline and of the corresponding outboard side be added to the pages. This precaution and the precautions stated above should assist in preventing confusion between Port and Starboard.

The fourth problem identified in the evaluation is that some operators were confused by the plant status model presented by the MCC. It is very important that the model of machinery control information presented by the MCC to the watchkeeper be as accurate and complete as possible. The matching of the watchkeeper's model and the MCC's model is mandatory if the benefits of digital systems are to be realized. Therefore, it has been recommended that the capabilities and limitations of the MCC be specifically explained during training. Also, any non-console actions which may be required should be indicated by a specific message on the appropriate page.

Discussion would be incomplete without comment on the lessons learned about the conduct of, and the requirements for, the evaluation of such complex, digital systems. Two major requirements were identified, one concerning the need to design the hardware and software to permit man-in-the-loop evaluation, the other concerned with the need for valid evaluation scenarios.

Electronic data-logging facilities must be included in the conceptual design of such prototype, or developmental systems. This is vital for performance assessment from the human engineering, training assessment, and incident analysis points of view. All models from initial prototype to operational versions require this capability if they are to be properly evaluated. Data-logging is imperative if important issues such as operator errors or slow operator responses which could imperil the ship are to be studied. Without the facility, it is not possible to thoroughly evaluate the system nor to design an optimal training scheme.

Recommendations (6) have been made as to the minimum type of data-logging necessary for the SMCC System. The simulator should keep a temporal record of all key depressions and control inputs made by the operator, together with a record of all system state variables. The criterion for adequate data-logging is that sufficient information must be stored in order to recreate the behaviour of the screens and computer graphic indicators. The recreation must then be useable to generate more detailed logs of information displayed on the screens as required. The logging system should have provision for a tape recorder with a minimum of two channels linked to the timing information for recording verbal exchanges.

Joystick values should be sampled at a rate of not less than twice a second and, preferably, provision should be made for a faster rate of at least one hundred times a second for short periods to investigate actual control strategies. Lastly, provision should be made to dump data to a removable storage medium, and software designed so that a run of several hours can be reconstructed and analyzed.

Another important prerequisite for the evaluation of systems like the watchkeeper's MCC is the preparation and testing of valid scenarios. Experience shows this to be a manpower intensive and time consuming activity which demands expert operator input. The scenarios run in this evaluation were designed to meet a limited purpose. Future evaluations should take into account more varied operations and conditions. These should include:

a) start up and slipping;

b) coming alongside and docking;

c) normal cruising and manoe..ring in a variety of sea states;

d) tactical manoeuvres in a variety of sea states;

e) replenishment at sea;

f) towing;

g) a variety of machinery faults, emergencies, and cascading alarms; and

h) battle damage in the control room such as dead CRT screens, damaged internal communications systems.

Such extensive exercising should confirm that no situation occurs where the system or some operators cannot cope.

CONCLUSION

The SMCC System is a first generation digital system which implements computer graphics at a watchkeeper's MCC to replace conventional process control instrumentation. The system appears to be viable within the total concept of ship machinery control. It demonstrates that machinery control information can be presented in a timely and readable format to meet the needs of watchkeepers. This can be achieved without inducing adverse levels of operator workload. The system evaluation confirmed the fact that sampling information on plant status through CRT screens rather than conventional instrumentation is an acceptable and adequate mode of operation.

The lessons learned from this case study are that long term planning of evaluations and development of detailed scenarios, and electronic data logging facilities will be required to refine and further advance this type of technology in the future.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the contributions made to this evaluation by DMEE 7 staff, Dr. Gaines of the Centre for Person-Computer Studies Inc., Dr. Moray of Man-Machine Systems Consultants Inc., and the Operators.

REFERENCES

(1) LCdr J.F. Carruthers, "SHINPADS A Ship Integration Concept", Proceedings of the FIFTH Ship Control Systems Symposium, Vol. 1, David W. Taylor Naval Research and Development Center, Annapolis, MD, USA, 1978.

(2) E.L. Gorrell, "Human Engineering Considerations in Design of Information Displays for Machinery Control Consoles of Future Naval Ships", Proceedings of the FIFTH Ship Control Systems Symposium, Vol. 4, David W. Taylor Naval Research and Development Center, Annapolis, MD, USA, 1978.

(3) Cdr J.E. Ironside, "SHINPADS An Integration Philosophy for the 21st Century", Proceedings of the SIXTH Ship Control Systems Symposium, Vol. 1, National Defence Headquarters, Ottawa, Ontario, Canada, 1981.

(4) E.L. Gorrell, "SHINMACS Machinery Control Console Design", Proceedings of the SIXTH Ship Control Systems Symposium, Vol. 4, National Defence Headquarters, Ottawa, Ontario, Canada, 1981.

(5) Cdr B.H. Baxter, LCdr R.J. Rhodenizer and P.V. Penny, "Shipboard Integrated Machinery Control System (SHINMACS) - A Canadian Forces Concept", Proceedings of the SIXTH Ship Control Systems Symposium, Vol. 5, National Defence Headquarters, Ottawa, Ontario, Canada, 1981.

(6) B.R. Gaines and N. Moray, "Human Engineering Evaluation of Standard Machinery Control Console System", SHINMACS-HEE-FIN:MAR 84.

(7) E.L. Gorrell, "Human Engineering Design Requirements for SHINMACS Machinery Control Consoles: Part II. Console Hardware and Functions", DCIEM Technical Report No. 80-R-27 (for Defence Purposes Only), Downsview, Ontario, Canada, 1980.

(8) E.L. Gorrell, "Human Engineering Design Requirements for SHINMACS Machinery Control Consoles: Part III. Console Displays and Operation", DCIEM Report No. 81-R-18, Downsview, Ontario, Canada, 1981.

(9) E.L. Gorrell, "SHINMACS Machinery Control Console Design", DCIEM Report No. 81-R-27, Downsview, Ontario, Canada, 1981.

(10) Cymbol Cybernetics Corporation, Standard Machinery Control Console User Manual (Baseline), SMCC - 4 - 1.0 Baseline, 12 December 1983.

(11) LCdr D.J. Marshall and LCdr P.J. MacGillivray, "A Modern Machinery Control Console For Warships", Proceedings of the SEVENTH Ship Control Systems Symposium, Bath, United Kingdom, 1984.