Chemical Engineering Journal 128 (2007) 33–46

Chemical Engineering Journal

www.elsevier.com/locate/cej

The design of water-using systems in petroleum refining using a water-pinch decomposition

Alberto Alva-Argáez [a], Antonis C. Kokossis [b,*], Robin Smith [c]

[a] CETC-Varennes, Natural Resources Canada, 1615 Lionel Boulet Blvd., Varennes Que., Canada

[b] Process & Information Systems Engineering, School of Engineering, University of Surrey, Guildford, Surrey, UK

[c] The University of Manchester, Manchester, UK

Received 15 March 2006; received in revised form 15 September 2006; accepted 2 October 2006

Abstract

Water reuse and recycling offer substantial potential for savings in petroleum refining, as the water volumes processed are large. Presently, there is a lack of methods to systematically screen and analyze design alternatives using a total systems approach. Such an approach would consider effluent treatment, recycle of treated water and freshwater distribution simultaneously. The paper contributes with a systematic methodology that empowers conceptual engineering and water-pinch [R. Smith, Chemical Process Design and Integration, 2nd ed., John Wiley & Sons, 2005; Y.P. Wang, R. Smith, Wastewater minimization, Chem. Eng. Sci. 49 (7) (1994) 981–1006; Y.P. Wang, R. Smith, Design of distributed effluent treatment systems, Chem. Eng. Sci. 49 (18) (1994) 3127–3145.] with mathematical programming methods. The method focuses on petroleum refineries explaining trade-offs and savings between freshwater costs, wastewater treatment, piping costs and environmental constraints on the discharge.

Keywords: Water management; Design; Optimization; Petroleum refining

1. Introduction

The total amount of water used in refineries has been estimated to an average 65–90 gallons of water per barrel of crude oil [7]. The waste effluents from petroleum refineries typically require treatment before reuse or discharge. Stringent regulations on the discharge are likely to become stricter, with restrictions applying not only to industrial users, but also to municipal wastewater treatment operations. Industry faces a challenge to reduce the wastewater it generates and attain sustainable standards of operation. Designs should incorporate economical solutions that address effluent segregation systems and the regeneration of effluents. The concentration of wastewater pollutants depends on the amount of process steams, as well as the amount and the composition of the process and cooling water in the plant. Water reuse and recycling define the final concentrations of pollutants. The standard practice is to bring together contaminated wastewater into a single wastewater stream [7,37] and treat it further in a central treatment facility. The practice produces streams that are more difficult to treat and invariably leads to higher processing costs.

There is an apparent lack of methods to systematically screen and analyze design alternatives with a total systems approach. Such an approach would consider effluent treatment, recycle of treated water and freshwater distribution simultaneously. The concept of mass exchange networks [8–11,18,32] should be attributed the first systematic approach to the problem. A methodology to design an optimal water network has been presented in [2]. Koppol et al. [25] investigated the impact of pursuing zero-discharge policies. MINLP formulations have been previously documented in the literature [14] to report challenges to handle the non-linear elements of the formulations. Early work to tackle bilinear terms includes applications of recursive LP and successive linear programming (SLP) [19–22,31,44]. Successful applications have been published in [3]. More systematic approaches to handle non-linearities are found in algorithms that employ mathematical decompositions based on generic schemes [1,4,12,13,15,36,39]. Examples include Branch and Bound strategies, cutting plane methods and, more recently genetic algorithms applied to small problems [35].

* Corresponding author. Tel.: +44 1483 686 573; fax: +44 1483 686 581.
E-mail address: a.kokossis@surrey.ac.uk (A.C. Kokossis).

1385-8947/$ – see front matter
doi:10.1016/j.cej.2006.10.001

Water-pinch is now an established concept in the literature. It is a targeting approach first developed by Wang and Smith [40,41] and later improved by a number of researchers [2,17,18,26–29,38] for applications in water re-use and wastewater minimisation. Water-pinch introduced the limiting water profile in graphical representations based upon concentration and mass loads. For each water consumer, water-pinch uses the maximum allowable inlet concentration, maximum allowable outlet concentrations and the mass load of each pollutant to be removed (or the water flow rate). A limiting composite curve can be constructed by combining all individual profiles into a single composite curve. Matching a water supply line provides targets for minimum freshwater consumption ahead of design. A manual design procedure has been developed to achieve targets.

A novel decomposition approach is advocated here that simplifies the challenges of the optimization problem, making systematic use of water-pinch insights to define successive projections in the solution space. The work builds on the propositions by Alva-Argaez et al. [2] and Gunaratnam et al. [17]. The approach employs a strategy to address mixed-integer linear programming formulations where binary variables enable the handling of network complexity and the imposition of practical constraints to ensure meaningful solutions. The formulation essentially extends the capabilities of the already powerful water-pinch analysis into multiple contaminants, accounting for capital costs in the objective, and the study of trade-offs between freshwater usage, mass exchanger costs, and the pipe network installation. The optimization directly accounts for cost elements and overcomes inherent limitations in multiple contaminants. The paper concludes with a generic methodology for the water system design of petroleum refineries and an approach with a potential to address retrofit problems.

2. Water users in a petroleum refinery

Refineries employ a wide spectrum of solvents (and contact solvent processes) with differential solubilities to extract desirable and undesirable feedstock components. The processes produce wastewater and are invariably designed to optimize the use, re-use and the recycles associated with each individual solvent. Oil and solvents are the main pollutants in the operation whereas the bottoms of the fractionation towers represent the major source of wastewater. Water users in refineries are widely reviewed in the literature [5,7,11,16,30,33,34,37,42] and summarized for the purposes of this study in Table 1.

In crude desalting the crude oil is dehydrated using a combination of emulsion-breaking and coalescence. The amount of water present in the crude oil is approximately 0.1–2 vol.% and the salts contained in the emulsified aqueous phase range from 10 to 250 pounds per thousand barrels (p.t.b). Salts are separated from oil either in the presence of chemicals (followed by heating and gravity separation) or under the influence of high voltage electrostatic fields that agglomerate the dispersed droplets. A single-stage electrical desalter can reduce the salt content down to 3 p.t.b. A two-stage unit can reduce it down to 0.3 p.t.b. Dehydration down to about 0.2 vol.% water on crude can usually be achieved. Make-up wash water is always added to the crude oil to assist the desalting process and it is used in the range of 3–10 vol.% on crude.

In crude oil distillation the wastewater comes from: (i) overhead accumulators (prior to recirculation or transfer from hydrocarbons to other fractionators) containing sulfides, ammonia, phenols, oil, chlorides and mercaptans; (ii) oil sampling lines; (iii) stable emulsions from barometric condensers used to create vacuum conditions (as barometric condensers are replaced by surface condensers, the oil vapours are not in contact with water); (iv) the overhead reflux drum where stripping steam is condensed with naphtha and contains H_2S and NH_3 (mostly as NH_4SH); (v) the overhead product drum that contains H_2S and possibly phenols. In thermal cracking, the major source of wastewater is the overhead accumulator on the fractionator, where water is separated from the hydrocarbon vapour and sent to the sewer system. Effluent streams can be expected to contain H_2S, NH_3 and phenols. Catalytic cracking units, are some of the largest sources of sour and phenolic wastewaters in a refinery. Pollutants from catalytic cracking generally come from the steam strippers and overhead accumulators on fractionators used to recover and separate the various hydrocarbon fractions produced in the cracking process. The major pollutants are oil, sulfides, phenols, cyanides, and ammonia. These pollutants produce an alkaline wastewater with high BOD and COD levels.

Hydrocracking is basically catalytic cracking in the presence of hydrogen and at least one wastewater stream from the process should be high in sulfides, since hydrocraking reduces the sulfur content of the material being cracked. Most of the sulfides are

Table 1
Major water users in petroleum refining

Water users	Water source	Wastewater main pollutants	Comments
Crude desalting	Fresh or foul water stripped	Free oil, ammonia, sulfides and suspended solids	3–10 vol.% on crude charge
Crude oil distillation	Stripping steam	Sulfides, ammonia, phenols, oil, chlorides, mercaptans	Oil sampling lines produce significant flows
Thermal cracking	Overhead accumulators	H_2S, ammonia, phenols	
Catalytic cracking	Stripping steam	Oil, sulfides, phenols, cyanide, ammonia	Alkaline wastewaters
Hydrocracking	Stripping steam	High in sulfides	
Polymerization	Pretreatment for H_2S removal	Sulfides, mercaptans, ammonia	Low volume but high strength waste
Alkylation	Caustic/water wash	Spent caustic, oil, sulfides	
Isomerization	Low demand	Low level of phenols	No sulfides, ammonia present
Reforming	Low demand	sulfide	Relatively clean process
Hydrotreating	Stripping steam	Ammonia, sulfides, phenol	Effluent depends heavily on feedstock

in the gas products, which are sent to a treating unit for removal and/or recovery of sulfur and ammonia. In polymerization, even though the process makes use of an acid catalyst, the waste is alkaline because the catalyst is recycled and any remaining acid is removed by caustic washing. Most of the waste comes from the pretreatment stage. The wastewater is high in sulfides, mercaptans and ammonia. Catalytic polymerisation plants may require disposal facilities for spent caustic containing sodium phosphate. In alkylation the major discharge is the spent caustics from the neutralisation stage. These wastewaters contain suspended and dissolved solids, sulfides, oils and other contaminants. Water drawn off from the overhead accumulators contains varying amounts of oil, sulfides and other contaminants but it is not a major source of waste.

Isomerization wastewaters present no major pollutant discharge problems. Sulfides and ammonia are not likely to be present in the effluent. Low levels of phenols and oxygen demand also should be expected. Reforming is a relatively clean process. The volume of wastewater flow is small and none of the waste streams has high concentrations of significant pollutants. The wastewater is alkaline and the major pollutant is sulfide from the overhead accumulator on the stripping tower used to remove light hydrocarbon fractions from the reactor effluent. Principal hydrotreating processes used include the pretreatment of catalytic reformer feedstock, naphtha desulfurization, lube oil polishing, pretreatment of cat cracking feedstock, heavy gas oil and residual desulfurization and naphtha saturation. The strength and quantity of the wastewater depends on the process used and the feedstock. Ammonia and sulfides are the primary contaminants but phenols may also be present. The catalytic hydrotreating of diesel oil to reduce its sulfur content does not per se produce sour water. However, the subsequent steam stripping of the hydrotreated diesel to remove the free H_2S does yield a sour condensate overhead along with a small amount of by product sour naphtha.

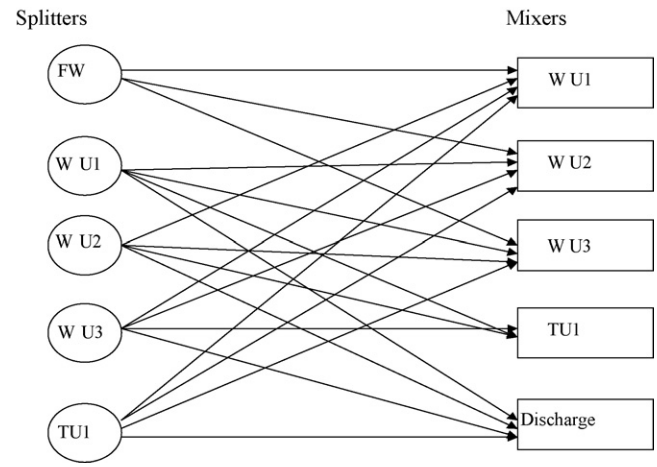

Fig. 1. Schematic representation of the superstructure model (single water source FW, water users WU1, WU2 and WU3 and treatment unit TU1. A single discharge point is illustrated).

3. Water network synthesis in a petroleum refinery

A design method is proposed that is based on a superstructure model formulated as a mixed-integer non-linear programming (MINLP) problem where every water user and treatment unit are represented by an inlet mixer and an outlet splitter; freshwater sources are introduced as splitters and wastewater discharge points as final mixers (Fig. 1). The non-linearities in the formulation are due to bilinear terms that appear in the mass balances (superstructure mixers and splitters), the non-linear terms of the sizing equations, and the cost functions for the water-using operations. All water users and treatment units have been modelled following the concepts developed by Wang and Smith in the water-pinch methodology [40,41].

3.1. Outline of the solution procedure

The approach integrates water-pinch with mathematical programming. Its overall iterative solution procedure is illustrated in Fig. 2. Let $(P_1)^k$ and $(P_2)^k$ be the formulation of the superstructure model with all bilinear terms projected onto the concentration space (MILP) and the set of mass balance equations projected onto the flowrate space (LP), respectively at iteration k, the algorithm involves the following steps:

1. Set $k=1$. Solve problem $(P_1)^k$ with $C_{c,i}^{\text{out},k} = C_{c,i}^{\max,\text{out},k}$ and $C_{c,t}^{\text{out},k} = 0$ to obtain z_1^{k*} and the optimal values of all the flow rates in the network, e.g. $F_i^{\text{eff,byp}^*}$, $F_t^{\text{out}^*}$, $F_{t,t'}^{it^*}$ and $F_{i,t}^{tr^*}$. Let $F_i^{\text{eff,byp},k} = F_i^{\text{eff,byp}^*}$, $F_t^{\text{out},k} = F_t^{\text{out}^*}$, $F_{t,t'}^{it,k} = F_{t,t'}^{it^*}$ and $F_{i,t}^{tr,k} = F_{i,t}^{tr^*}$.
2. Solve problem $(P_2)^k$ for a new vector of outlet concentrations, $C_{c,t}^{\text{out},*}$. Set $C_{c,t}^{\text{out},k+1} = C_{c,t}^{\text{out},*}$.
3. Set $k=k+1$, $p^{k+1}=p^k \times 10$ (penalty value in objective function), and solve problem $(P_1)^{k+1}$ to obtain new values of the flow rates in the network. Let the optimal objective value be z_1^{k+1}.
4. If $\chi^{\inf \cdot k} \leq \varepsilon_2$ or $|z_{\text{opt}}^k - z_{\text{opt}}^{k+1}/z_{\text{opt}}^k| \leq \varepsilon_1$, stop. Otherwise, go to step 2.
5. Solve the NLP that results from problem $(P_2)^k$ with all structural features fixed according to the solution from step 4.

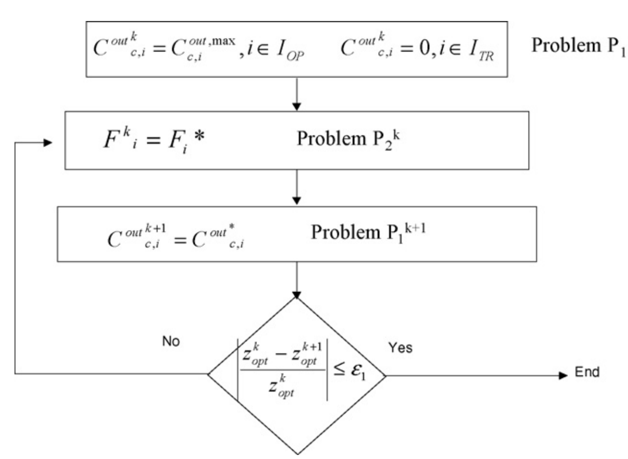

Fig. 2. Iterative procedure to find a feasible solution.

Table 2
Data for the process streams involved in mass exchangers (case study)

Process	Contaminant	$C^{in,PS}$ (ppm)	$C^{out,PS}$ (ppm)	Flow rate (t/h)	Mass load (g/h)
HDS	HC	–	–	54.222	3,400
	H_2S	8000	350		414,800
	Salt	–	–		4,590
Desalter	HC	–	–	263.15	800
	H_2S	–	–		200
	Salt	4000	200		1,000,000

The method iterates between the two problems:

- the change in the objective function of $(P_1)^k$ less than a tolerance, ε_1 (typically around 5% change), and
- feasibility is reached, i.e., the term $\chi^{\mathrm{inf},k}$ is sufficiently small (tolerance ε_2).

The solution stands as an initial point for an NLP optimization of the un-projected solution. The implementation of the algorithm relied on the GAMS modelling environment [6], OSL was the solver selected for all MILP and LP formulations and CONOPT was employed for the NLP stage. In order to specify the vector of outlet concentrations in all water users, the approach builds on the insights from water-pinch methodologies where all contaminants tend to reach their maximum limits and, in the optimal solution, at least one contaminant will be at its maximum limit. The initial assumption of perfect treatment performance defines an infeasible point for the original problem (P), as no treatment unit can have $RR_{c,t} = 1$. To account for the infeasibilties, a penalty function is introduced in the objective function of (P_1), which drives the distance from the feasible region to zero. The uncoupling of the bilinear terms (flow times concentration) with the introduction of the mass flows and the relaxation-projection strategy allows the search to be performed in the space defined by the convex feasible region of problem $(P_1)^k$. The choice of weights for the χ^{inf} term follows the suggestions made in [31]. In all the examples, the alternative assignment of p-values as preset multiples (usually 10^3 times) of the largest of the Lagrange multipliers associated with active constraints [39,44] yields premature convergence.

The approach is generic and applies for an arbitrary set of contaminants. The majority of problems involve: hydrocarbons (HC), hydrogen sulfide (H_2S) and salts (salts). Water users are usually among:

- atmospheric distillation unit (CDU);
- vacuum distillation operation (VDU);
- hydrotreating (HDS);
- crude desalting (desalter);
- cooling tower (C. tower);
- boiler house;
- delayed coker;
- other operations grouped together (others).

The steam stripping operation of HDS and the desalting operation are water-using operations classified as mass exchange units. The information on the streams processed with water in the mass exchange operations is given in Table 2 (key contaminants). An alternative multi-contaminant approach requires only additional information and sizing equations in the models. For mass exchangers, the number of equilibrium stages are calculated for each contaminant ($N_{c,i}$). The exchanger is sized according to the contaminant calling for the largest number of stages using:

$$N_i \geq N_{c,i} \quad \forall c \in C$$

Vapour–liquid equilibria used in Kremser type of relations [24] are derived from data available from a simulation study. For the HDS and the Desalter these are assumed to be linear. More specifically, for the HDS steam stripping (key contaminant: hydrocarbons) these are:

$$y^* = 0.6121x$$

whereas for the desalter (key contaminant is salts) this is:

$$y^* = 0.0018x$$

In both expressions, y^* is the equilibrium concentration in the process stream and x is the corresponding concentration in the water stream (in ppm). The economic criteria are applied as shown in Table 3.

The cost functions for the stripping column are given by:

$$\mathrm{cost}^{\mathrm{shell}}(\$) = 10,240H$$
$$\mathrm{cost}^{\mathrm{tray}}(\$/\mathrm{tray}) = 490e^{0.8D}$$

H and D indicate the column height and diameter (in m). The cost of the trays is approximated by a piecewise linear expression of the form:

$$D = \sum_q D_q^F \lambda_q$$
$$\mathrm{cost}^{\mathrm{tray}} = \sum_q \mathrm{cost}_q^F \lambda_q \quad \text{for } q = 1, 2, \ldots, N_q$$

The set q is used to define the number of linear segments. For the column diameter, a range between 0.5 and 4 m is considered

Table 3
Economic criteria

Freshwater cost	0.3\$/t
Operating time	8600 h/year
Interest rate	15%
Time of return on investment (leading to an annualisation factor of 0.438)	3 years
Installation factor for piping cost	4

Table 4
Matrix of distances for the oil refinery

Source	Sink								
	S1	O1	O2	O3	O4	O5	O6	O7	Discharge
S1	0	25	225	50	40	35	200	250	1000
O1	25	0	300	50	75	450	100	150	600
O2	225	300	0	60	100	65	225	250	500
O3	50	50	60	0	250	150	75	350	400
O4	40	75	100	250	0	350	175	100	325
O5	35	100	65	150	350	0	125	75	150
O6	200	100	125	75	175	125	0	275	125
07	250	150	250	350	100	75	275	0	50

(i.e. $D_1^F = 0.5$ and $D_{N_q}^F = 4$) with two linear segments (i.e., $N_q = 3$). The number of real trays is calculated using a 60% overall column efficiency. The desalting cost is expressed by:

(i) A fixed capital cost for each desalting stage of $722,895 and
(ii) An operating cost of $612,262/year per stage (to account for electricity and chemicals usage) [16].

Cost calculations further assume:

- Stainless steel construction;
- 20 min residence time;
- Horizontal cylindrical tank;
- $L/D = 3$.

The distances (in m) between operations are given in Table 4. Carbon steel piping is assumed as a standard and the cost of the pipework is calculated by [23]:

$$\text{Cost}^{\text{pipe}} = 3,603.4 \cdot A + 124.46\ [\$/\text{m length}]$$

3.1.1. Base case

The operating data is given in Table 5 and the total water use is 1250 t/h. Even though the base case does not consider optimization, it still accounts for some water reuse, as the effluent from the CDU steam stripping feeds the single-stage desalter. A fixed water loss of 770 t/h is assumed for the evaporative cooling system; the remaining flow is available for reuse and treatment. The flow in the cooling system can be reduced to 813.3 t/h without implications on the heat removal of the plant. The costs associated with the base case are presented as follows:

- Column O2: (10 trays) Total cost of $150,454, annualised cost of $65,898/year.

Table 5
Base case design

Operation	Flow (t/h)	Contaminant	C^{in} (ppm)	C^{out} (ppm)	Load (g/h)
CDU	50.00	HC	0.00	13.50	675.00
		H_2S	0.00	360.00	18000.00
		Salt	0.00	31.50	1575.00
HDS	48.00	HC	0.00	70.83	3400.00
		H_2S	0.00	8641.67	414800.00
		Salt	0.00	95.63	4590.00
Desalter	10.00	HC	13.50	93.63	801.28
		H_2S	360.00	380.03	200.32
		Salt	31.50	100031.50	1000000.00
Others	25.00	HC	0.00	16.72	418.00
		H_2S	0.00	91.20	2280.00
		Salt	0.00	22.80	570.00
Cooling tower	900.00	HC	0.00	10.83	9750.00
		H_2S	0.00	1.44	1300.00
		Salt	0.00	14.44	13000.00
Boiler house	190.00	HC	0.00	0.00	0.00
		H_2S	0.00	0.00	0.00
		Salt	0.00	1789.47	340000.00
Delayed coker	37.00	HC	0.00	133.24	4930.00
		H_2S	0.00	2727.57	100920.00
		Salt	0.00	156.76	5800.00

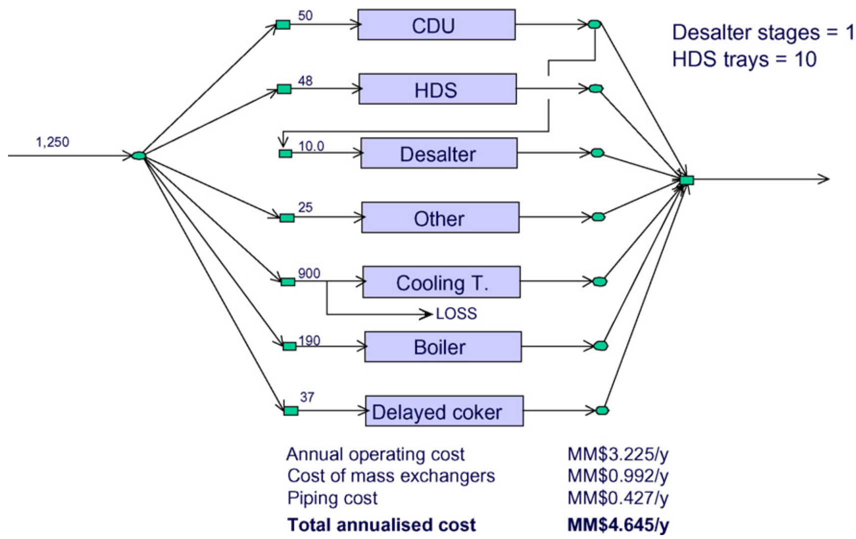

Fig. 3. Base case design.

- Desalter O3: (single-stage) Total cost of $722,896, annualised cost of $316,629 plus annual operating cost of the unit of $612,262/year. In total, $928,891/year.
- Water cost is $3,225,000/year.
- Pipework cost (assumed flow velocity, 2 m/s everywhere) $426,412/year.

The total annual cost of the base case is $4,645,202/year. The limiting water profile is obtained through a combination of equilibrium considerations (key contaminants only) and considerations of corrosion and fouling [5,26–29,33]. Table 5 presents the limiting water profile data. The outlet concentrations in the base case are lower than the maximum limits identified in limiting data set, opening up opportunities to improve water efficiency.

The design of the base case is shown in Fig. 3 and uses a single-stage desalting unit. The stripping column in HDS has 10 trays for the separation.

3.1.2. Analysis

A number of scenarios are explored to illustrate the importance of the systems approach. The variants produced consider different objectives and illustrate the potential of the approach to systematically produce optimal solutions. The problem is first solved to minimize the water consumption without and with maximum reuse (cases A and B, respectively). Next, the objective is augmented to consider capital cost associated with mass exchangers (neglecting piping and layout costs) (case C). Finally, a design is produced that addresses all design aspects (case D).

3.1.2.1. Case A. The optimal design is produced and shown in Fig. 4. The case reduces the overall freshwater consumption by enabling the water streams to reach the maximum possible concentration in each operation. Within the specifications given by Table 6, it is no longer possible to satisfy the requirements of the desalter purely with reused water from CDU.

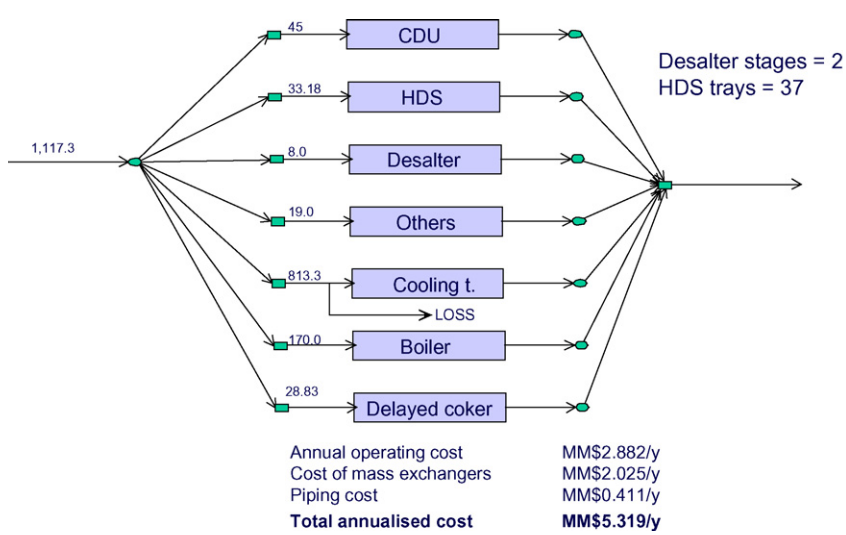

Fig. 4. Design for case A.

Table 6
Limiting water profile data for the water-using operations

Process	Contaminant	$C^{\text{in,max}}$ (ppm)	$C^{\text{out,max}}$ (ppm)	Mass load (g/h)
CDU	HC	0	15	675
	H_2S	0	400	18,000
	Salt	0	35	1,575
HDS	HC	20	120	3,400
	H_2S	300	125,00	414,800
	Salt	45	180	4,590
Desalter	HC	120	220	801.2
	H_2S	20	45	200.32
	Salt	200	125,000	1,000,000
Others	HC	0	22	418
	H_2S	0	120	2,280
	Salt	0	30	570
Cooling tower	HC	150	225	9,750
	H_2S	200	310	1,300
	Salt	250	350	13,000
Boiler house	HC	0	0	800
	H_2S	0	0	200
	Salt	0	2,000	340,000
Delayed coker	HC	100	270	4,930
	H_2S	20	3,500	100,920
	Salt	50	250	5,800

The outlet concentration of H_2S from CDU is too high to use in the desalter. The freshwater consumption is subsequently reduced via a once-through policy with no water re-use in place.

The water consumption is reduced by 10%. However, the implications in the capital cost of the mass exchangers offset water savings. Even though the water is used more efficiently, the resulting design is expensive. The cost of the desalter increases significantly as it changes from a single-stage unit to a two-stage operation. It is interesting to note the operation is not as sensitive to the mass transfer driving forces (due to the high solubility of the salts in the water stream) as it is to the actual water flow rate through the unit. Thus, there is no motivation to use freshwater for desalting and increase the flow to perform the separation in a single stage is a desirable option.

3.1.2.2. Case B. The optimization problem is solved with the automated method outlined above and based on the data of Table 5. The objective function is the minimum freshwater cost. The procedure converges in three iterations. The primal problem features 770 continuous variables and the formulation accounts for 72 binary variables. The solution consumes 54.1 CPU s overall. The NLP consumes 0.71 CPU s to converge to the solution in Fig. 5.

The freshwater demand is reduced by around 15% against the base case. This is due to the maximization of the water

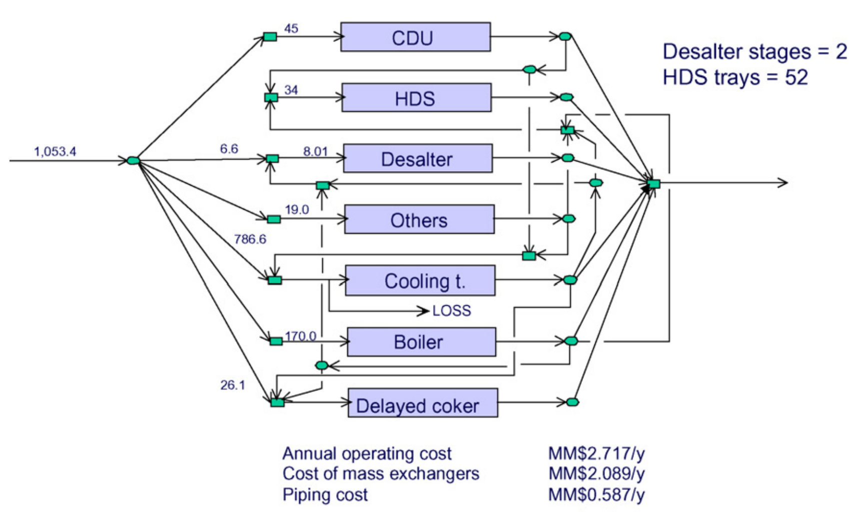

Fig. 5. Design for case B.

re-use. However, the network structure is pretty complex with some of the flows being relatively small (less than 1 t/h). Constraints can be introduced to eliminate small streams as a manual manipulation of the solution. A more systematic approach is left for discussion later (case D). Alongside the freshwater reduction, an increase in the capital cost is observed and the design emerges more expensive than the once-through scenario. This is attributed to the increase in the number of stages that are selected in the mass exchangers. In the once-through case, there have been 37 trays for the HDS and two for the desalting stages. In this case the number of trays is increased to 52 for HDS and there are two stages for desalting. The increase has subsequently affected the capital cost of the stripping operation. The increased number of connections in the design should be translated into higher capital cost piping. Indeed, the latter increased by 30% as against the once-through case.

Cases A and B illustrate the fact that a design approach addressing the maximum reuse (or, equivalently, the minimization of water) may lead to rather complex designs. Although such designs can meet the targets, they are not practical or functional solutions for implementation. Combined with a general observation that the water target can be met by a multitude of nearby solutions, we step into a cost-based objective function (cases C and D) letting comparisons with the previous cases unfold from the analysis.

3.1.2.3. Case C. The objective function is formulated as the minimum total annual cost, where the cost of mass exchangers is considered. The solution is presented in Fig. 6.

The procedure required four iterations to converge consuming 32.1 CPU s. Primal problems featured 784 continuous and 72 binary variables. The NLP required 11.4 CPU s.

The major feature in this solution is that the cost of the mass exchangers is reduced by almost 50%. Looking at the inlets to the two mass exchange operations it can be observed that the solution provides for larger flows of water. The number of stages in HDS is reduced to ten trays (compare with 52 in the minimum freshwater cost design). The desalting operation is performed in a single stage. Since the desalting stages are major capital items, the reduction in the cost is significant, whereas the increase in the freshwater cost has been marginal (about 3% increase). The network structure still remains complex though, as characterized by the presence of small flows. Again, it is possible to eliminate the small streams as a manual manipulation of the solution.

The freshwater supply to the cooling tower is increased. The effluent though improves in quality and can be re-used into the mass exchange operations. Compare this situation with the re-use pattern obtained in the minimum freshwater cost design (Fig. 4) where the cooling tower is mainly a sink for reused water. The flow of reused water from the cooling tower to the mass exchangers is increased in the last design and achieves a lower capital cost. This design reduces freshwater as compared to the base case by 13.5%; the cost of the mass exchangers is increased by 4% as mass transfer driving forces in the steam stripping operation (HDS) are reduced. The total annual cost is 2.5% less but the piping cost has increased its share by more than 100%.

3.1.2.4. Case D. The objective is the minimum annual cost and considers explicitly freshwater, mass exchangers, and piping costs. The design obtained is shown in Fig. 7.

The procedure converged in three iterations and consumed 36.1 CPU s. Primal problems comprised 792 continuous and 72 binary variables. The NLP takes 11.86 CPU s to converge. This final design shows the result when the complete picture is considered in the optimization procedure. The capital cost is similar with case C but the structure of the network is greatly simplified with the piping cost accordingly reduced. Ten trays are used in the stripping column (HDS), and a single desalting stage. The piping cost is reduced by more than 50% in comparison with case C. A summary of the results is presented in Table 7 and in Fig. 8.

One should note the existence of a significant number of network configurations that are able to operate at the same water volume (major drive of the operating costs), but each featur-

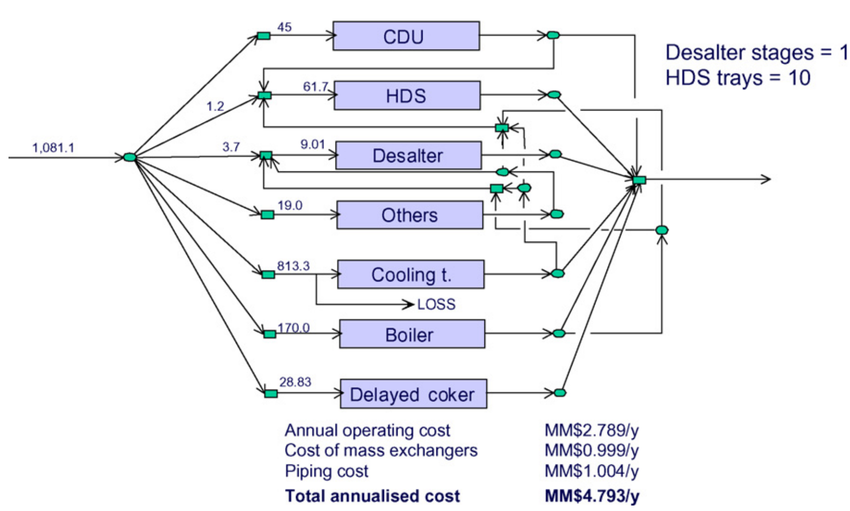

Fig. 6. Design for case C.

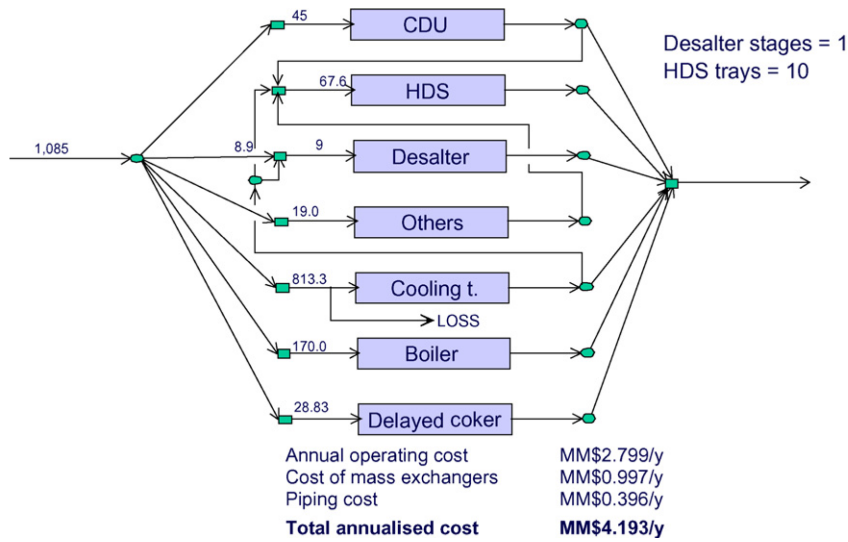

Fig. 7. Design for case D.

Table 7
Comparison of results

Case	Operating cost (MM$/year)	Mass exchangers cost (MM$/year)	Piping cost (MM$/year)	TAC (MM$/year)
Base case	3.23	0.99	0.21	4.42
A	2.88	2.02	0.20	5.09
B	2.72	2.08	0.28	5.08
C	2.79	1.04	0.48	4.31
D	2.80	1.04	0.20	4.04

ing very different capital costs. Most notably, the size of mass exchanger units may change with the different flow patterns as a result of the different patterns for mass transfer. In the presented solutions, the desalting operation represents a major capital cost and the addition of additional stages implies a significant capital investment. In the case of the stripping column, the number of stages is less crucial.

The method can be used to explore trade-offs between operating and capital cost in the water system. As compared with the base case, the final solution achieves total annual savings of approximately $400,000 (or 8.7%). Coupled with these overall savings, freshwater consumption can be reduced by some 13%, that is 165 t/h. The calculation of the trays was found sensitive to the equilibrium parameters. The estimation of these parameters is accordingly expected to have a significant bearing in the results.

3.2. Sensitivity analysis

A number of sensitivity experiments are presented next. The effect of an increase in the freshwater cost is illustrated in Fig. 9 where the total annualised cost (in MM$/year) is plotted as a function of the cost for the freshwater supply (in $/t).

The linear relationship is explained in that the selected networks are variants of similar structures, featuring only minor

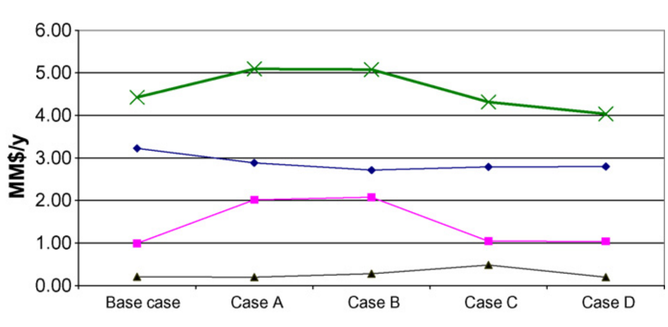

Fig. 8. Comparison of results.

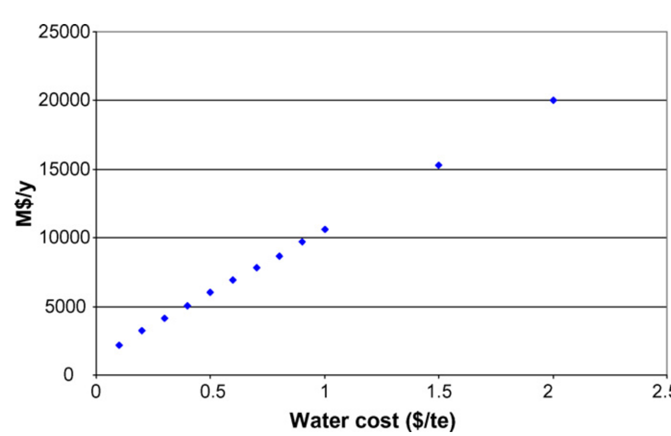

Fig. 9. Sensitivity of the design to freshwater costs.

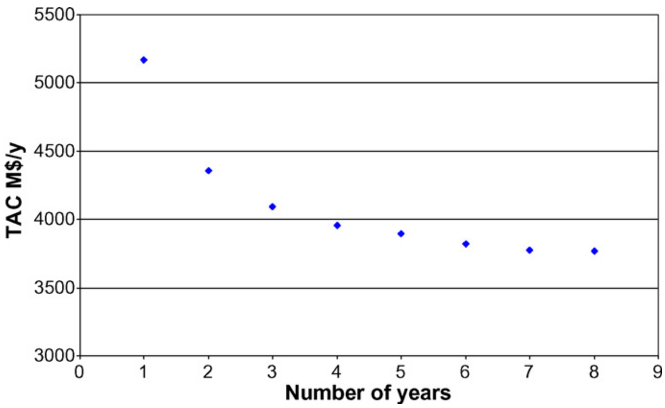

Fig. 10. Total cost of the design vs. return on investment (time).

changes in the flows involved. A review of freshwater intake as a function of cost demonstrates that the demanded flow remains essentially constant for a range of prices up to $2/te. Similarly, the behaviour with respect to the capital cost of the network reveals that the cost of mass exchangers features minor variations and the design appears, within the range of values analysed, robust to freshwater costs. The desalter requires a single stage. Freshwater costs should exceed $100/t before a second stage is justified. Savings of freshwater – achieved with additional stages in stripping operation – are marginal and the stripper size remains at 10 trays throughout the range of the analysis. Note that in the water network of Fig. 7 (minimum freshwater cost), the stripping operation is fed by reused water only. The stream from "Others" is consumed completely in the HDS stripper but the stream from CDU has some spare capacity.

The sensitivity of the design is finally assessed with respect to the different economic criteria employed. Designs assumed 3 years annualisation and a rate of return at 15%. Fig. 10 presents results considering different time horizons for capital return.

The shape of the plot explains that the water targets are robust to annualisation parameters. The freshwater consumption for the designs implied in Fig. 10 indicate designs processing similar volumes of freshwater. The underlying network structures are similar for year 1–4 and change from year 5.

The structural changes that appear in year 5 yield an alternative configuration. The designs obtained for year 4 and 5 are

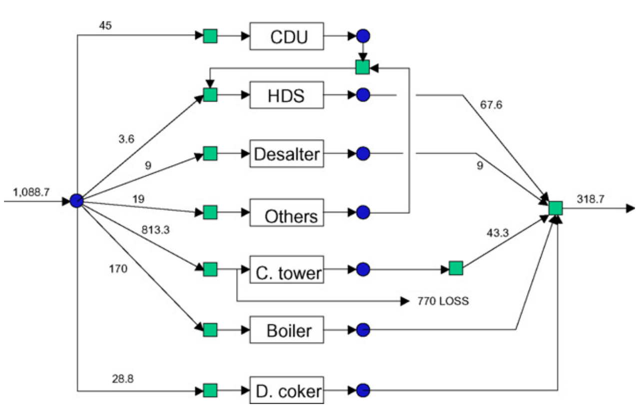

Fig. 11. Four years (Design I).

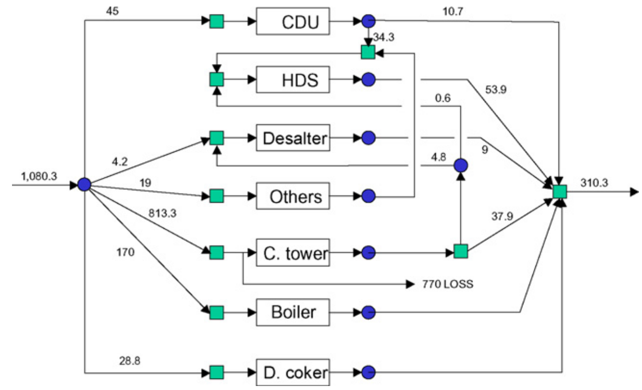

Fig. 12. Five years (Design II).

Table 8
Comparison of designs

	I	II	Difference (%)
Total cost (M$/year)	4042.50	4152.82	−2.73
Freshwater cost (M$/year)	2809.02	2787.36	0.77
Piping cost (M$/year)	516.66	648.68	−25.55
ME cost (M$/year)	716.82	716.78	0.01

shown in Figs. 11 and 12, respectively. Their differences are constrained in the number of re-use connections. The design for year 4 (Design I) years features only two re-use connections (from CDU and others into HDS), whereas the design in year 5 (Design II) year shows similar connections with the addition of some reuse out of the cooling tower and into desalter and the HDS unit. Design I consumes slightly more water than Design II and is offset by savings in capital costs (mainly lower piping costs due to fewer reuse connections). Table 8 shows an overall comparison of these results.

4. Retrofit applications

The extension of the decomposition approach is next illustrated with retrofit applications. Most apparently, it is assumed that an existing system (units and pipe network) is already installed. In addition to the grassroots design, the model is faced with a challenge to reduce structural and operational modifications. The objective is accordingly adjusted to search for the retrofit options where the optimal use of existing capital is made.

The following modifications were required to the model.

4.1. Pipe network

Based on the installed network, bounds are established on the flow that can potentially pass through the pipes (expressed as functions of their sizes). The existence of a pipe remains a degree of freedom and a zero cost is assigned for the use of existing pipes. Smaller or larger pipes accordingly yield penalties in the objective as they appear as "new investment" to make. In order to achieve this, a new set of constraints is introduced in the mathematical formulation. The use of special ordered sets [6,43] aided the modeling without a need to increase the number of binary variables. The new constraints are introduced as follows:

$$A^{\mathrm{IP}}_{i,i'} = \sum_{l} A^{l}_{i,i'}\sigma_l \quad (1)$$

$$\mathrm{cost}^{\mathrm{IP,pipe}}_{i,i'} = \sum_{l} \mathrm{cost}^{\mathrm{IP,pipe},l}_{i,i'}\sigma_l \quad (2)$$

$$\sum_{l} \sigma_l = 1 \quad (3)$$

where the set $L = \{l/l\ is\ a$ bound for a linear segment$\}$ defines the degree of linearisation. $A^{\mathrm{IP}}_{i,i'}$ represents the cross-sectional area of a pipe connecting operations i to i'. The $\sigma_l(1,2,\ldots,NL)$ are SOS2 variables. An alternative formulation based on binary variables is described next. Let δ be a 0–1 variable and let:

$$A^{\mathrm{IP}^D}_{i,i'} = A^{\mathrm{IP}}_{i,i'} \cdot \delta_{i,i'} \quad (4)$$

$$A^{\mathrm{IP}}_{i,i'} - A^{\mathrm{IPUe}}_{i,i'}\delta_{i,i'} \leq 0 \quad (5)$$

$$A^{\mathrm{IP}}_{i,i'} - UA^{\mathrm{IP}}_{i,i'}\delta_{i,i'} \leq A^{\mathrm{IPUe}}_{i,i'} \quad (6)$$

Namely,

$$A^{\mathrm{IPUe}}_{i,i'} \leq A^{\mathrm{IP}}_{i,i'} \leq UA^{\mathrm{IP}}_{i,i'} \quad \text{then} \quad \delta_{i,i'} = 1 \quad \text{and} \quad \delta_{i,i'} = 0 \text{ otherwise} \quad (7)$$

$$\mathrm{cost}^{\mathrm{IP,pipe}}_{i,i'} = a^{\mathrm{IP,pipe}}_{i,i'} A^{\mathrm{IPD}}_{i,i'} + b^{\mathrm{IP,pipe}}_{i,i'}\delta_{i,i'} \quad (8)$$

$$A^{\mathrm{IP}}_{i,i'} - UA^{\mathrm{IP}}_{i,i'}(1-\delta_{i,i'}) \leq A^{\mathrm{IPD}}_{i,i'} \quad (9)$$

$$A^{\mathrm{IPD}}_{i,i'} \leq A^{\mathrm{IP}}_{i,i'} \quad (10)$$

$$A^{\mathrm{IPD}}_{i,i'} \leq UA^{\mathrm{IP}}_{i,i'} \cdot \delta_{i,i'} \quad (11)$$

$A^{\mathrm{IP},l}_{i,i'}$ are the bounds on the linear segments for the cross-sectional area of the existing pipe (i,i'). $A^{IPD}_{i,i'}$ in (4) represent bilinear terms with binary variables. Constraints (4)–(6) ensure that if the cross sectional area of the pipe connecting operation i to operation i' exceeds the existing size, $A^{\mathrm{IPUe}}_{i,i'}$, then $\delta_{i,i'} = 1 UA^{\mathrm{IP}}_{i,i'}$ is the upper bound on the total flow of the connection. Once the need for a new pipe is identified, the cost function from Eq. (8) becomes linear in $A^{\mathrm{IPD}}_{i,i'}$. Constraints (9)–(11) are introduced for each existing connection.

4.2. Unit costs

These include capital costs related to the mass exchanger units. Concerning modifications required for a staged unit (e.g. stripping column), a fixed column diameter is used as a reference (existing column) letting retrofit options consider changes in the number of trays. A solution featuring additional equilibrium stages could be translated as a recommended modification to the internals of the column (e.g. use of high efficiency trays), additional trays, or the addition of a new column. The additional capital cost is evaluated considering the cost of replacement and modifications. As in the previous cases, a piecewise linear representation that is based on SOS2 sets is employed:

$$N_i = \sum_{h} N^{h}_{i}\omega_h \quad (12)$$

$$\mathrm{cost}^{\mathrm{ME}}_{i} = \sum_{h} \mathrm{cost}^{\mathrm{ME},h}_{i}\omega_h \quad (13)$$

$$\sum_{h} \omega_h = 1 \quad (14)$$

And alternatively, the variable β_i [0,1] is introduced to identify the number of stages in the solution. Let $N'_i = N_i\beta_i$ and:

$$N_i - (N^{E}_{i}E^{0}_{i})\beta_i \geq 0 \quad (15)$$

$$N_i = N^{U}_{i}\beta_i \leq (N^{E}_{i}E^{0}) \quad (16)$$

$$\text{if} \quad (N^{E}_{i}E^{0}_{i}) \leq N_i \leq N^{U}_{i}, \quad \text{then,} \quad \beta_i = 1 \quad \text{and} \quad \beta_i = 0 \text{ otherwise} \quad (17)$$

$$\mathrm{cost}^{\mathrm{ME}}_{i} = N'_i CF_i \quad (18)$$

$$N_i - N^{U}_{i}(1-\beta_i) \leq N'_i \quad (19)$$

$$N'_i = N_i \quad (20)$$

$$N'_i \leq N^{U}_{i}\beta_i \quad (21)$$

$$N'_i \geq 0 \quad (22)$$

Constraints (12)–(14) model the existing column. The set $H = \{h|h$ is a bound for a linear segment$\}$ defines the degree of linearisation. Parameters N^{h}_{i} represent the assumed range of number of equilibrium stages; ω_h are defined as SOS2 variables. The parameters $\mathrm{cost}^{\mathrm{ME},h}_{i}$ in (13) correspond to cost values for the range of equilibrium stages considered. The formulation defined by constraints (15)–(22) involves binary variables. Parameters N^{E}_{i} represent the existing number of equilibrium stages. The lower bound is set to zero. Constraints (15) and (16) are required equilibrium stages with respect to N^{E}_{i}. The cost of the column is defined by introducing a new variable N' (18). Constraints (19)–(22) complete the formulation.

5. Retrofit applications. Case study

The problem of Section 3 is selected with the existing connections listed in Table 9 (together with their maximum flows). A simple structure is assumed for the existing system with reuse connection between CDU and the Desalter. Upper bounds $A^{\mathrm{IPUe}}_{i,i'}$, are added to the mathematical formulation for the existing connections and their cost set to zero. From the base case, the existing stripping column has 10 trays and the diameter is 2.9 m. The parameters associated with the cost of the mass exchangers in Eq. (18) are given by:

$$\mathrm{CF} = \left(\mathrm{cost}^{\mathrm{tray,HE}}\left(\frac{E^{0}}{E^{0,\mathrm{HE}}}\right)\right) \quad (23)$$

Table 9
Existing connections and maximum flows for retrofit application

Source	Sink								
	S1	O1	O2	O3	O4	O5	O6	O7	Discharge
S1	0	50	48	0	25	900	190	37	0
O1	0	0	0	10	0	0	0	0	40
O2	0	0	0	0	0	0	0	0	48
O3	0	0	0	0	0	0	0	0	10
O4	0	0	0	0	0	0	0	0	25
O5	0	0	0	0	0	0	0	0	130
O6	0	0	0	0	0	0	0	0	190
O7	0	0	0	0	0	0	0	0	37

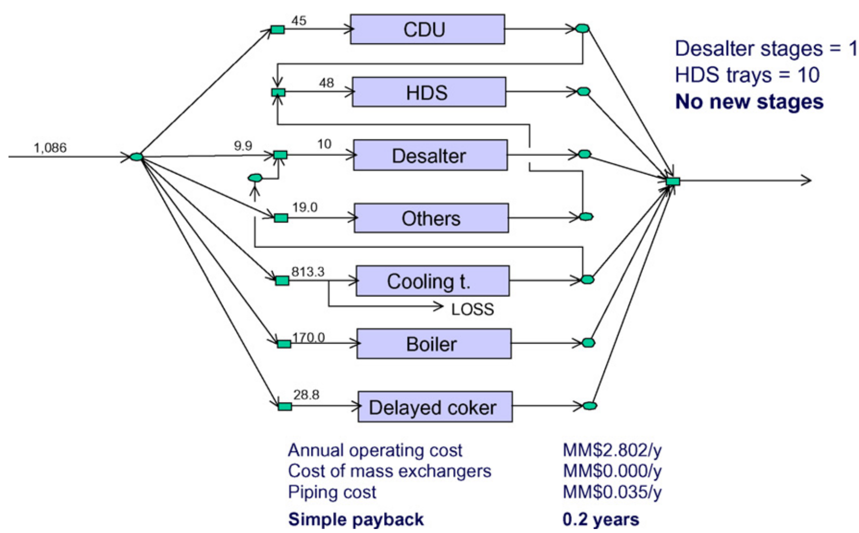

Fig. 13. Retrofit design.

where E^0 is the overall column efficiency; ε_2 the tolerance value; $E^{0,\mathrm{HE}}$the overall column efficiency using high efficiency trays; $\mathrm{cost}^{\mathrm{tray,HE}}$: cost of each high efficiency tray for the given diameter.

The desalter is modelled in a simpler way (cost factor $\mathrm{CF}_{\mathrm{DES}}$) as it can operate only as a single stage (the only possibility is to add another unit). The option of not using desalter is excluded and the cost function is modified to:

$$C_{\mathrm{DES}} = (N - 1)\mathrm{CF}_{\mathrm{DES}} \tag{24}$$

The above expression yields a zero cost if a single stage is required adding penalties if a second desalting stage is selected. The retrofit problem is solved using the proposed approach and the selected design is illustrated in Fig. 13. Primal problems feature 834 continuous variables and 72 binaries. The solution algorithm requires four iterations to converge consuming 101.9 CPU s. The MINLP model features 823 continuous variables and 85 binary variables consuming 211.1 CPU s to converge. The costs associated with this design are presented in Table 10.

The freshwater consumption is reduced by 13% compared to the base case. The solution represents annual savings in freshwater costs of $423,120. The capital investment required for pipes amounts to $35,183/year. The total capital investment is $80,331 leading a payback period of 0.2 years (or approximately 3 months). According to the results presented, the pipes to install correspond to the following connections:

Source	Sink
S1	Desalter
CDU	HDS
Others	HDS
Cooling tower	Desalter

Existing pipes can be used for the remaining connections. The mass exchangers remain unchanged.

The method has produced results with relative ease considering the actual time spent in developing such scenarios manually for industrial projects. Solutions apparently depend upon the initial solutions but the retrofit studies carried out so far indicate that the computational burden produced is much lighter than the one accounted in the grassroots designs.

Table 10
Retrofit design economic data

Freshwater cost [$/year]	2,802,086.45
Mass exchangers [$/year]	0
Piping cost [$/year]	35,183
Total annual cost [$/year]	2,837,269.4

6. Conclusions

The paper presents a systematic approach to address water reuse in oil refineries. The methodology is based on the water-pinch decomposition and addresses a mixed-integer non-linear programming formulation that features integer variables for the connections between units. The model classifies and models major water users in refineries demonstrating its potential to address both grassroots and retrofit problems. Studies report reductions of freshwater consumption by over 10% and with minimal capital investment. The approach is further used to study the sensitivity of the solution on a variety of design and economic parameters.

The approach further enables the analysis of alternative networks useful in practical stages where several scenarios are required to screen and preview. Results indicate there is a multitude of similar solutions one needs to screen systematically. As retrofit options, the approach enables the installation of new pipes in the network, the replacement of stripping column internals with the introduction of high efficiency trays. Designs demonstrate the exhaustive search enabled by the approach. The approach is fully automated and computationally inexpensive. The design engineer can thus experiment with different scenarios, different objective functions, and can assess trade-offs between reuse performance and cost.

Acknowledgments

The financial support of CONACYT, EPSRC and the Process Integration Department at UMIST is gratefully acknowledged. The support from CETC-Varennes, Natural Resources Canada for the production of the paper is also acknowledged.

References

[1] F.A. Al-Khayyal, J.E. Falk, Jointly constrained biconvex programming, Math. Oper. Res. 8 (2) (1983) 273–286.

[2] Alva-Argaez, A. (1999) "Integrated design of water systems," PhD Thesis Department of Process Integration, UMIST, Manchester, UK, p. 199.

[3] T.E. Baker, L.S. Lasdon, Succesive linear programming at Exxon", Manage. Sci. 31 (1985) 264–274.

[4] M.S. Bazaraa, H.D. Sherali, C.M. Shetty, Nonlinear Programming: Theory and Algorithms, 2nd ed., John Wiley and Sons Inc., New York, 1993.

[5] M.R. Beychok, Aqueous Wastes from Petroleum and Petrochemical Plants, 1st ed., John Wiley and Sons Inc., London, 1967.

[6] A. Brooke, D. Kendrick, A. Meeraus, GAMS, a User's Guide, The Scientific Press, San Francisco, CA, 1988.

[7] W. Byers, W. Doerr, R. Krishnan, D. Peters, How to Implement Industrial Water Reuse. A Systematic Approach, vol. 1, AIChE, New York, 1995.

[8] M.M. El-Halwagi, V. Manousiouthakis, Synthesis of mass-exchange networks, AIChE. J. 8 (1989) 1233–1244.

[9] M.M. El-Halwagi, V. Manousiouthakis, Automatic synthesis of mass-exchange networks with single component targets, Chem. Eng. Sci. 9 (1990) 2813–2831.

[10] M.M. El-Halwagi, V. Manousiouthakis, Simultaneous synthesis of mass-exchange and regeneration networks, AIChE. J. 36 (1990) 1209–1219.

[11] M.M. El-Halwagi, A.M. El-Halwagi, V. Manousiouthakis, Optimal design of dephenolization networks for petroleum refinery wastes, Trans. IChemE 70 (Part B) (1992) 131–139.

[12] C.A. Floudas, A. Aggarwal, A.R. Ciric, Global optimum search for nonconvex NLP and MINLP problems, Comp. Chem. Eng. 13 (10) (1989) 1117–1132.

[13] C.A. Floudas, V. Visweswaran, A global optimization algorithm (GOP) for certain classes of nonconvex NLPs-I. Theory', Comp. Chem. Eng. 14 (12) (1990) 1397–1417.

[14] C.A. Floudas, Nonlinear and Mixed-Integer Optimization, 1st ed., Oxford University Press, New York, 1995.

[15] B. Galan, I.E. Grossman, Optimal design of distributed wastewater treatment networks, Ind. Eng. Chem. Res. 37 (10) (1998) 4036–4048.

[16] J.H. Gary, G.E. Handwerk, Petroleum Refining: Technology and Economics, 3rd ed., Marcel Dekker, New York USA, 1994.

[17] M. Gunaratnam, A. Alva-Argáez, A.C. Kokossis, J.-K. Kim, R. Smith, Automated design of total water systems, Ind. Eng. Chem. Res. 44 (2005) 588–599.

[18] N. Hallale, D.M. Fraser, Capital cost targets for mass exchange networks. A special case: water minimisation, Chem. Eng. Sci. 53 (2) (1998) 293–313.

[19] C.A. Haverly, Recursion model behaviour. More studies, ACM Sigmap Bull. 28 (1980) 39–41.

[20] C.A. Haverly, Behaviour of LP models, ACM Sigmap Bull. 22 (1977) 52–60.

[21] C.A. Haverly, Studies of the behaviour of recursion for the pooling problem, ACM Sigmap Bull. 25 (1978) 19–29.

[22] C.A. Haverly, Behaviour of recursion model More studies, ACM Sigmap Bull. 26 (1979) 22–28.

[23] IChemE, A guide to capital cost estimating, 3rd ed., IChemE and The Association of Cost Engineers, London, 1988.

[24] C.J. King, Separation Processes, 2nd ed., McGraw-Hill, Inc., New York, NY, 1980.

[25] A.P.R. Koppol, M. Bagajewicz, B.J. Dericks, M.J. Savelski, On zero water discharge solutions in the process industry, Adv. Environ. Res. 8 (2003) 151–171.

[26] Kuo, W. J. 1996. "A combined approach to water minimisation and effluent treatment system design", Ph.D. dissertation. Department of Process Integration. UMIST, Manchester, UK.

[27] W.J. Kuo, R. Smith, Effluent treatment system design, Chem. Eng. Sci. 52 (23) (1997) 4273–4290.

[28] W.J. Kuo, R. Smith, Designing for the interactions between water-use and effluent treatment, Trans. IChemE 76 (Part A) (1998) 287–301.

[29] W.J. Kuo, R. Smith, Design of water-using systems involving regeneration, Trans. IChemE 76 (Part B) (1998) 94–114.

[30] Metcalf, Eddy, Wastewater Engineering; Treatment Disposal and Reuse, 3rd ed., Mc Graw Hill, New York, 1991.

[31] F. Palacios-Gomez, L. Lasdon, M. Engquist, Nonlinear optimization by successive linear programming", Man. Sci. 28 (10) (1982) 1106–1120.

[32] K.P. Papalexandri, E.N. Pistikopoulos, A. Floudas, Mass exchange networks for waste minimization: a simultaneous approach, Trans. IChemE 72 (Part A) (1994) 279–294.

[33] Perez Huerta, S. 1995. "Capital and operating cost trade-offs in wastewater minimisation". MSc Dissertation, Process Integration Department, UMIST, Manchester.

[34] J.W. Porter, J.H. Blake, R.T. Milligan, Zero discharge of wastewater from petroleum refineries, in: Complete Water Reuse: Industry's opportunity. Based on papers presented at the National Conference on Complete Water Reuse, April 23–27, 1973, pp. 448–457.

[35] D. Prakoptol, T. Srinophakun, GApinch: genetic algorithm toolbox for water pinch technology, Chem. Eng. Process. 43 (2004) 203–217.

[36] I. Quesada, I.E. Grossman, Global optimization of bilinear process networks with multicomponent flows, Comp. Chem. Eng. 19 (12) (1995) 1219–1242.

[37] M. Sittig, Petroleum Refining Industry Energy Savings and Environmental Control, 1st ed., Noyes Data Corporation, New York, 1978.

[38] R. Smith, Chemical Process Design and Integration, 2nd ed., John Wiley & Sons, 2005.

[39] J. Viswanathan, I.E. Grossman, A combined penalty function and outer-approximation method for MINLP optimization, Comp. Chem. Eng. 14 (7) (1990) 769–782.

[40] Y.P. Wang, R. Smith, Wastewater minimization, Chem. Eng. Sci. 49 (7) (1994) 981–1006.
[41] Y.P. Wang, R. Smith, Design of distributed effluent treatment systems, Chem. Eng. Sci. 49 (18) (1994) 3127–3145.
[42] E. Weisberg, D.L. Stockton, Water reuse in a petroleum refinery complete water reuse: industry's opportunity, in: Based on Papers Presented at the National Conference on Complete Water Reuse, April 23–27, 1973, pp. 400–406.
[43] H.P. Williams, Model Building in Mathematical Programming, 3rd ed., John Wiley and Sons, Surrey, UK, 1997.
[44] J. Zhang, N.H. Kim, L. Lasdon, An improved successive linear programming algorithm, Man. Sci. 31 (10) (1985) 1312–1331.